EVOLUTIONARY AND ECOLOGICAL GENOMICS IN DEEP-SEA ORGANISMS

By

Santiago Herrera Monroy

B.Sc., Universidad de los Andes, Bogotá, Colombia, 2007
M.Sc., Universidad de los Andes, Bogotá, Colombia, 2009

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

and the

WOODS HOLE OCEANOGRAPHIC INSTITUTION

February 2015

Signature of Author …….
Joint Program in Oceanography/Applied Ocean Science and Engineering
Massachusetts Institute of Technology and Woods Hole Oceanographic Institution
December 30, 2014

Certified by …….
Dr. Timothy M. Shank
Thesis Supervisor

Accepted by …….
Professor Martin F. Polz
Chair, Joint Committee for Biological Oceanography
Massachusetts Institute of Technology and Woods Hole Oceanographic Institution

EVOLUTIONARY AND ECOLOGICAL GENOMICS IN DEEP-SEA ORGANISMS

By Santiago Herrera Monroy

Submitted to the MIT/WHOI Joint Program in Oceanography/Applied Ocean Science and Engineering
on December 30, 2014 in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Biological Oceanography

ABSTRACT

Hydrothermal vents and coral ecosystems are conspicuous biological hot spots in the deep-sea. These ecosystems face increasing threats from human activities. Having thorough taxonomic inventories as well as understanding species' relatedness, genetic diversity, connectivity patterns, and adaptive potential is fundamental for the implementation of conservation strategies that help mitigate these threats. This thesis provides fundamental high-priority knowledge in taxonomic, evolutionary, and ecological aspects of deep-sea coral and vent species, by harnessing the power of genomic tools and overcoming long-standing methodological barriers. First, I develop bioinformatic tools that help guide the design of studies aiming to characterize eukaryotic genome diversity using restriction-site associated DNA sequencing. Using these tools I find that the predictability of restriction site frequencies in eukaryotic genomes is chiefly determined by the phylogenetic position of the target species and the recognition sequence of the selected restriction enzyme. These tools are then applied to test global-scale historical biogeographic hypotheses of vent fauna using barnacles as model. Phylogeographic inferences suggest that the western Pacific was the place of origin of the major vent barnacle lineage, followed by circumglobal colonization eastward along the southern hemisphere during the Neogene. I suggest that the geological processes and dispersal mechanisms discussed here can explain distribution patterns of many other marine taxa in addition to barnacles. Regional-scale analyses indicate that vent barnacle populations are well connected within basins and ridge systems, and that their diversity patterns do not conform to the predictions from the hypothesis that seamounts are centers of endemism. I then move on to resolve long-standing questions regarding species definitions in recalcitrant deep-sea coral taxa, by unambiguously resolving evolutionary relationships and objectively inferring species boundaries. Finally, I explore the adaptive potential of deep-sea coral species to environmental changes by examining a case of adaptation to shallow water from the deep sea. Candidate positive-selection markers shared between pairs of shallow and deep populations are identified as likely makers for genomic regions involved in adaptation. Overall, the results from this thesis constitute critical baseline data with which to assess potential effects of anthropogenic disturbances on deep-sea ecosystems.

Thesis supervisor:
Dr. Timothy M. Shank
Title: Associate Scientist with Tenure, Biology Department, Woods Hole Oceanographic Institution

Para mis papás, Liliana y Mauricio

AKNOWLEDGEMENTS

This degree is the fulfillment of a lifelong dream. I am infinitely grateful to Tim Shank, for taking a chance with me, for challenging me constantly, for honest advising, for providing limitless opportunities to pursue my interests and develop my skills, for unconditional support and trust, and for making me feel more like a colleague than a student since the first day. I could not imagine a better place to have done my PhD. I look forward to many more ventures to come.

I am extremely grateful to my committee Ann Tarrant, Adam Reitzel, and Martin Polz for guiding me along the way and being my best advocates. Ann, thank you for always having an open door, for being ready to help at a minute´s notice, and for being generous sharing your knowledge and giving thorough feedback. Adam, thank you for your contagious enthusiasm for science and inspiring creativity; I greatly enjoy working with you. I am also grateful for Mark Hahn for chairing my defense and offering excellent advise.

To Juan Sánchez, Steve Cairns, and Allen Collins, thank you for your mentorship and support, which fostered my passion for science and lead me to pursue this degree.

Special thanks to the Academic Programs Office at WHOI, the MIT-WHOI Joint Program Office, the MIT/WHOI Joint Program Committee, the Joint Committee for Biological Oceanography, the administrative staff at both WHOI and MIT, and the support, maintenance, security and facilities staff at WHOI for your dedication and hard work behind the scenes. I am also thankful to all the course instructors and teaching assistance for sharing their knowledge and fostering critical thinking.

To all my friends in the lab and around Woods Hole, particularly Ellie Bors, Arthur Olive, Cat Munro, Larry George, Jason Kapit, Kalina Gospodinova, Amalia Aruda, Albert Almada, Kristen Hunter-Cevera, Kerry McCulloch, Taylor Heyl, Annette Govindarajan, Walter Cho, Breea Govenar, Kate Buckman, Taylor Crockford, Liz Drenkard and Britta Voss, thank you for making this such an enjoyable time.

Gracias a toda mi familia por su amor y apoyo, pero en especial a mis padres Mauricio y Liliana. Papás, todos mis logros se los debo a ustedes. Gracias por su incondicionalidad, sacrificio, y ejemplo.

To Jill, thank you for being my loyal partner through this journey, and I look forward to being by your side in all the journeys to come.

I thank P. Aldersdale (CSIRO), N. Ardila (ECOMAR) and J. Sanchez (U. Andes) for assistance with morphological specimen identifications. I also thank my outstanding diving team E. O'Brien (WHOI), D. Forsman (WHOI), J. Fellows (WHOI), and Canadian collaborators V. Tunnicliffe (U. Victoria), J. & S. Schooner, K. Heylar, and N. McDaniel for invaluable assistance during scuba diving fieldwork in British Columbia (DFO scientific license FIN130270). The times spent underwater are some of my favorite memories.

This research was supported by the Office of Ocean Exploration and Research of the National Oceanic and Atmospheric Administration (NA09OAR4320129 to TMS); the Division of Ocean Sciences of the National Science Foundation (OCE-1131620 to TMS); the Division of Polar Programs of the National Science Foundation (PLR-0739675 to TMS); the Astrobiology Science and Technology for Exploring Planets program of the National Aeronautics and Space Administration (NNX09AB76G to TMS); the National Geographic Society/Waitt Foundation (W285-13 to SH); and the Academic Programs Office (Ocean Ventures Fund and fellowship support to SH), the Ocean Exploration Institute (fellowship support

to TMS and SH) and the Ocean Life Institute (internal grant to TMS and SH) of the Woods Hole Oceanographic Institution.

For enabling access to key specimens I thank S. Mills (NIWA), K. Schnabel (NIWA), D. Tracey (NIWA), M. Clark (NIWA), A. Rowden (NIWA), K. Iizasa (U. Tokyo), Y. Suzuki (U. Tokyo), S. Nakagawa (JAMSTEC), P. Tyler (NOCS), J. Copley (NOCS), A. Rogers (Oxford), N. Roterman (Oxford), J. Resing (NOAA-PMEL), R. Embley (NOAA-PMEL), A. Reysenbach (PSU), M.K. Tivey (WHOI), P. Fryer (UH), C. Langmuir (Harvard), K. Von Damm (UNH), M. Lilley (UW), S. Cairns (Smithsonian), E. Cordes (Temple U.), A. Quattrini (Temple U.), G. Workman (DFO), M. Wyeth (DFO), K. Anderson (DFO), M. Frey (RBCM), H. Gartner (RBCM), Ed Bowlby (NOAA), E. Edinger (Memorial U.), J. Sanchez (U. Andes), L. Watling (U. Hawaii), J. Adkins (CalTech) the masters, crew, scientific personnel, and funding agencies of expeditions AT03-28, AT07-06, AT07-35, JC042, JC067, KM0417, KM0912, KOK0505, KOK0506, Lophelia II 2009, NT97-10, NT97-14, NT99-09, RB-0503, RR1211, TAN1007, TAN1104, TAN1206, TAN1213, TN234, TN236, YK06-13, and YK09-13.

Specimens provided by the National Institute of Water and Atmospheric Research (NIWA) were collected under research programs: Kermadec Arc Minerals, funded by the New Zealand Ministry of Business, Innovation & Employment (MBIE), Auckland University, Institute of Geological and Nuclear Science (GNS), and WHOI; Ocean Survey 20/20 funded by Land Information New Zealand; Impact of resource use on vulnerable deep-sea communities (CO1X0906), funded by MBIE; Nascent Inter-Ridge Volcanic And Neotectonic Activity, funded by the Ministry for Primary Industries (MPI), GNS, MBIE, and the U. of New Hampshire; Scientific Observer Program funded by MPI; and the Joint New Zealand-USA 2005 NOAA Ring of Fire Expedition, part of NIWA's Seamount Program (FRST CO1X0508).

CONTENTS

CHAPTER 1

Introduction

Threats to deep-sea coral ecosystems

Deep-sea corals are some of the most conspicuous invertebrate inhabitants of hard-bottom deep-sea benthic environments worldwide. They are not only more diverse in terms of number of species than their shallow counterparts (Cairns 2007), but they also play a fundamental role as foundation species and ecosystem engineers, creating three-dimensional habitats that are occupied by a high diversity of associate species (Buhl-Mortensen & Mortensen 2005; Costello *et al.* 2005; Buhl-Mortensen *et al.* 2010; Watling *et al.* 2011). Deep-sea ecosystems also support fisheries (D'Onghia *et al.* 2011) and have been identified as important sources of marine natural products (Leal *et al.* 2012). Deep-sea corals, generally speaking, have evolved in a relatively stable and energy-poor environment. They are characterized by slow growth rates, extreme longevity, low fecundity, and a late age of maturity (Roberts *et al.* 2009). These characteristics make deep-sea coral ecosystems fragile and with low resilience to the severe disturbances generated by many modern human activities, including physical damage caused by bottom-trawling fishing (Watling & Norse 1998; Koslow *et al.* 2001; Waller *et al.* 2007; Althaus *et al.* 2009; Clark & Rowden 2009; Williams *et al.* 2010), climate change and ocean acidification caused by emissions of greenhouse gasses (Doney *et al.* 2009), pollution and habitat destruction generated by waste disposal (Kvassnes *et al.* 2012), deep-sea mining (Van Dover 2010), and off-shore drilling for hydrocarbons (White *et al.* 2012). As such, the United Nations has designated deep-sea coral ecosystems as Vulnerable Marine Ecosystems (http://www.un.org/depts/los/general_assembly/general_assembly_resolutions.htm UN General Assembly resolutions 61/105 and 64/72), requiring new management and protection strategies, such as the U.S. Magnuson-Stevens Fishery Conservation and Management Reauthorization Act of 2006. The Magnuson-Stevens act authorized Regional Fishery Management Councils to designate zones to protect deep-sea corals from damage caused by fishing gear, and established a Deep-Sea Coral Research and Technology Program to advance a critical understanding of their taxonomy and systematics, biogeography, genetic connectivity, and physiological responses to stressors in deep water coral ecosystems.

Threats to hydrothermal vent ecosystems

Benthic chemosynthetic ecosystems present a sharp contrast to other ecosystems in the deep-sea. They are characterized by high rates of *in situ* primary productivity, marked patchiness, and highly dynamic geological settings (Van Dover 2000). Deep-sea hydrothermal vent environments can have extremely

steep gradients of chemistry and temperature, and high disturbance frequencies given their occurrence on volcanic or actively spreading tectonic systems. As a consequence, hydrothermal vent environments present extreme selective pressures on evolutionary timescales, and can yield vent ecosystems with relatively low biodiversity and high endemism (Van Dover 2010). These very characteristics may make them susceptible to disturbances caused by mining of polymetallic sulfides. Although organisms at deep-sea hydrothermal vents have adapted to cope with natural disturbances, the frequency and magnitude at which these occur can vary greatly depending on the particular geophysical nature of each system (Baker & German 2004). Thus, disturbance from mining could have additive effects to natural disturbances at scales not previously experienced by these organisms, which could potentially lead to significant losses of biodiversity. There is a surging need to provide the highest-priority information needed to design optimal conservation and management strategies for areas that are being prospected for mining (Van Dover *et al.* 2012). These priorities emphasize the identification of conservation units at the genetic, species and biogeographic levels, and a better understanding of connectivity among populations.

Priorities for conservation of deep-sea ecosystems

Knowledge of conservation units is fundamental for the creation and implementation of efficient strategies that help mitigate the effects of harmful human activities on deep-sea ecosystems. Such knowledge must include well-founded taxonomic inventories that allow us to identify species and ecosystems at risk, as well as an understanding of their relatedness, genetic variance, distribution, connectivity patterns, and adaptation potential (Christensen *et al.* 1996; Dubois 2003; Roberts & Cairns 2014). Nevertheless, gaining this knowledge in deep-sea ecosystems is difficult due to the extreme challenges of working in these environments, combined with the paucity of genetic resources for deep-sea taxa.

Issues of traditional phylogenetic approaches

Traditionally, phylogenetic and population genetic studies in non-model organisms – which aims include understanding species boundaries, relationship patterns, evolutionary histories, factors that diminish or promote genetic diversity, demographic processes of populations, and interactions with environmental conditions – have based their power on a handful of homologous DNA sequence markers. Target DNA sequence markers can be easily sequenced using nearly-universal primers; however, several problems have been identified with the use of the few traditional sequence markers available for non-model organisms (e.g., mitochondrial and ribosomal genes), including low variability, biased loci sampling, poor genome representation and small statistical power, presence of pseudogenes, multiple gene copies, and non-independence caused by linkage (Brumfield 2003; Brito & Edwards 2008). Microsatellites emerged

during the last decade as a novel class of markers promising great potential to solve population-level questions. However, many problems have identified with the use of microsatellites, including poor understanding of their mutational processes, high rates of back mutations and homoplasy, presence of null alleles, low reproducibility and comparability of results, and high monetary and time expenses for individual marker development and genotyping (Brumfield 2003).

Opportunities of novel genomic approaches

The problems related to the use of traditional genetic markers have been recognized and accounted for in model organisms by comparing large amounts of genomic sequence information among individuals and identifying thousands of variable regions, such as single nucleotide polymorphisms (SNPs) across the genome, e.g. Clark *et al.* (2007) and Rokas *et al.* (2003). Single nucleotide polymorphisms (SNPs) have been recognized as the most prevalent source of variability in any given genome; they represent *ca.* 90% of the genetic variation in the human genome (Collins *et al.* 1999). As such, SNPs overcome most of the problems related with the use of traditional sequence-markers - they are present in extremely high numbers across the genome, have a wide range of mutational rates, behave largely as independent loci, and can be screened with high-throughput techniques, making them economical (Brumfield 2003). Technological and methodological developments in next-generation sequencing platforms over the last five years (e.g., Illumina, PacBio, IonTorrent, etc) have made genomic resources for SNP development and genotyping increasingly accessible and available to researchers investigating a wide spectrum of evolutionary questions in diverse organisms. Their use is now being successfully implemented in several non-model organisms, thus offering a great opportunity to overcome the difficulties inherent to the use of traditional approaches in many taxa.

SNPs have been successfully used to resolve evolutionary and biogeographic histories of diverse taxa, from bees (Whitfield *et al.* 2006), to humans (Jakobsson *et al.* 2008), nematodes (Andersen *et al.* 2012), anemones (Reitzel *et al.* 2013), and pitcher plant mosquitoes (Emerson *et al.* 2010). Recently developed methodologies allow the implementation of next-generation sequencing technologies for the rapid detection and genotyping of SNPs in organisms without reference genomes (Garvin *et al.* 2010); the so-called genotyping-by-sequencing (GBS) approaches (Baird *et al.* 2008). Restriction-site associated DNA sequencing (RAD-seq), a kind of GBS approach, enables high-throughput sequencing of homologous sites in nuclear genome after a complexity-reduction step carried out with a high-fidelity restriction enzyme. This revolutionary approach has opened a new frontier in molecular studies, with the promise of providing profound insights into the genetics, organismal biology, ecology and evolution of wild populations (Seeb *et al.* 2011).

Objective and outline

With this thesis, I aim to provide fundamental high-priority knowledge in taxonomic, genetic, evolutionary, and ecological aspects of deep-sea coral and hydrothermal vent species, by harnessing the power of novel genomic tools. This knowledge could be applied to the conservation and management of deep-sea ecosystems and their biodiversity.

To achieve this goal, I have performed bioinformatic and empirical studies implementing restriction-site associated DNA sequencing methodologies on ecologically important deep-sea coral and hydrothermal vent species. I present results from my investigations on the causes and consequences of evolutionary forces that determine biodiversity patterns in the deep-sea. In Chapter 2, I test the hypothesis that genome composition can be used to predict the number of restriction sites for a given combination of restriction enzyme and genome across the eukaryotic tree of life. I develop a methodology to predict the frequency of restriction sites that helps guide the design of GBS studies in eukaryotic organisms. This methodology is then used throughout the rest of the thesis. Chapters 3-6 can be categorized into chapters that examine macroevolutionary processes above the species level (chapters 3 and 5), and chapters that examine microevolutionary processes within species (chapters 4 and 6). Alternatively, they can also be categorized according to the target taxon system: deep-sea hydrothermal vent barnacles (chapters 3 and 4), or deep-sea corals (chapters 5 and 6) (Fig. 1).

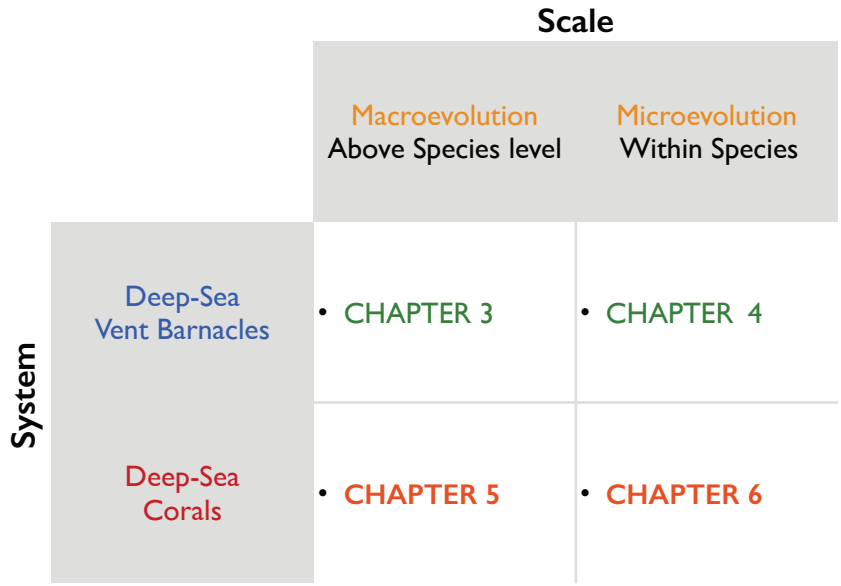


Figure 1. Conceptual categorization of chapters according to the time scale of the processes and the target taxon systems examined.

In Chapter 3, I compare traditional DNA sequence makers with novel genomic data from restriction site associated DNA sequencing (RAD-seq) to characterize the global genetic diversity of barnacles from deep-sea hydrothermal vents, which due to their widespread but discontinuous distribution represent an excellent model for testing global biogeographic hypotheses. I infer their time and place of origin, mode of dispersal, and diversification throughout the world's vents. I then complement this research in Chapter 4, by using vent barnacles as a model to test smaller-scale biodiversity hypothesis. Specifically, I test the hypothesis that seamounts act as islands promoting divergence and speciation in deep-sea vent fauna. For this I compare the genetic diversity contained in single nucleotide polymorphisms (SNPs) obtained through RAD-seq to examine population-structuring patterns of populations in different barnacle species from seamount and spreading ridges.

In Chapter 5, I demonstrate the empirical utility of RAD-seq to solve evolutionary questions in deeper time by unambiguously resolving phylogenetic relationships among recalcitrant octocoral taxa with divergences greater than 80 million years and performing unambiguous species delimitations. In Chapter 6, I make use of the knowledge gained in defining species boundaries in deep-sea coral species by addressing questions regarding interactions of coral populations from the same species with their environment, in shorter time scales. I focus on a deep-sea coral species that can also be found in shallow high-latitude fjords with distinct environmental conditions from those found in the deep-sea. Here I aim to identify and characterize genomic regions that have may have enabled the successful adaptation to shallow-water in this deep-sea species. Finally, I summarize findings and draw general conclusions in Chapter 7.

REFERENCES

Althaus F, Williams A, Schlacher T*, et al.* (2009) Impacts of bottom trawling on deep-coral ecosystems of seamounts are long-lasting. *Marine Ecology Progress Series* **397**, 279-294.

Andersen EC, Gerke JP, Shapiro JA*, et al.* (2012) Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nature Genetics* **44**, 285-U283.

Baird NA, Etter PD, Atwood TS*, et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**, e3376.

Baker ET, German CR (2004) On the global distribution of hydrothermal vent fields. In: *Mid-Ocean Ridges*, pp. 245-266. American Geophysical Union.

Brito PH, Edwards SV (2008) Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica* **135**, 439-455.

Brumfield R (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution* **18**, 249-256.

Buhl-Mortensen L, Mortensen PB (2005) Distribution and diversity of species associated with deep-sea gorgonian corals off Atlantic Canada. In: *Cold-Water Corals and Ecosystems* (eds. Freiwald AR, Roberts JM), pp. 849-879. Springer-Verlag, Berlin Heidelberg.

Buhl-Mortensen L, Vanreusel A, Gooday AJ, *et al.* (2010) Biological structures as a source of habitat heterogeneity and biodiversity on the deep ocean margins. *Marine Ecology* **31**, 21-50.

Cairns S (2007) Deep-water corals: an overview with special reference to diversity and distribution of deep-water scleractinian corals. *Bulletin of Marine Science* **81**, 311-322.

Christensen NL, Bartuska AM, Brown JH, *et al.* (1996) The report of the Ecological Society of America committee on the scientific basis for ecosystem management. *Ecological Applications* **6**, 665-691.

Clark AG, Eisen MB, Smith DR, *et al.* (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203-218.

Clark MR, Rowden AA (2009) Effect of deepwater trawling on the macro-invertebrate assemblages of seamounts on the Chatham Rise, New Zealand. *Deep-Sea Research Part I* **56**, 1540-1554.

Collins FS, Brooks LD, Chakravarti A (1999) A DNA polymorphism discovery resource for research on human genetic variation (vol 8, pg 1229, 1998). *Genome Research* **9**, 210-210.

Costello M, McCrea M, Freiwald A, *et al.* (2005) Role of cold-water *Lophelia pertusa* coral reefs as fish habitat in the NE Atlantic. In: *Cold-Water Corals and Ecosystems* (eds. Freiwald A, Roberts JM), pp. 771-805. Springer Berlin Heidelberg.

D'Onghia G, Indennidate A, Giove A, *et al.* (2011) Distribution and behaviour of deep-sea benthopelagic fauna observed using towed cameras in the Santa Maria di Leuca cold-water coral province. *Marine Ecology Progress Series* **443**, 95-110.

Doney SC, Balch WM, Fabry VJ, Feely RA (2009) Ocean acidification: A critical emerging problem for the ocean sciences. *Oceanography* **22**, 16–25.

Dubois A (2003) The relationships between taxonomy and conservation biology in the century of extinctions. *Comptes Rendus Biologies* **326**, S9-S21.

Emerson KJ, Merz CR, Catchen JM, *et al.* (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 16196-16200.

Garvin MR, Saitoh K, Gharrett AJ (2010) Application of single nucleotide polymorphisms to non-model species: a technical review. *Molecular Ecology Resources* **10**, 915-934.

Jakobsson M, Scholz SW, Scheet P, *et al.* (2008) Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* **451**, 998-1003.

Koslow JA, Gowlett-Holmes K, Lowry JK, *et al.* (2001) Seamount benthic macrofauna off southern Tasmania: community structure and impacts of trawling. *Marine Ecology Progress Series* **213**, 111-125.

Kvassnes AJ, Sweetman AK, Hobæk A, Thorseth IH, Bolam S (2012) Imptail – Improved submarine tailing placements (STPS) in Norwegian Fjords. In: *Ocean Sciences Meeting*. Abstract ID:11074, Salt Lake City, Utah.

Leal MC, Puga J, Serodio J, Gomes NCM, Calado R (2012) Trends in the discovery of new marine natural products from invertebrates over the last two decades - where and what are ee bioprospecting? *PLoS One* **7**, e30580.

Reitzel AM, Herrera S, Layden MJ, Martindale MQ, Shank TM (2013) Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular Ecology* **22**, 2953-2970.

Roberts JM, Cairns SD (2014) Cold-water corals in a changing ocean. *Current Opinion in Environmental Sustainability* **7**, 118-126.

Roberts JM, Wheeler A, Freiwald AR, Cairns SD (2009) *Cold-Water Corals : The Biology and Geology of Deep-Sea Coral Habitats* Cambridge University Press, Cambridge, UK ; New York.

Rokas A, Williams BL, King N, Carroll SB (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**, 798-804.

Seeb LW, Templin WD, Sato S, *et al.* (2011) Single nucleotide polymorphisms across a species' range: implications for conservation studies of Pacific salmon. *Molecular Ecology Resources* **11**, 195-217.

Van Dover CL (2000) *The Ecology of Deep-Sea Hydrothermal Vents* Princeton University Press, Princeton, N.J.

Van Dover CL (2010) Mining seafloor massive sulphides and biodiversity: what is at risk? *ICES Journal of Marine Science* **68**, 341-348.

Van Dover CL, Smith CR, Ardron J*, et al.* (2012) Designating networks of chemosynthetic ecosystem reserves in the deep sea. *Marine Policy* **36**, 378-381.

Waller R, Watling L, Auster P, Shank T (2007) Anthropogenic impacts on the Corner Rise seamounts north-west Atlantic Ocean. *Journal of the Marine Biological Association of the United Kingdom* **87**, 1075-1076.

Watling L, France SC, Pante E, Simpson A (2011) Biology of deep-water octocorals. *Advances in Marine Biology* **60**, 41-122.

Watling L, Norse EA (1998) Disturbance of the seabed by Mobile Fishing Gear: A Comparison to Forest Clearcutting. *Conservation Biology* **12**, 1180-1197.

White HK, Hsing P-Y, Cho W*, et al.* (2012) Impact of the Deepwater Horizon oil spill on a deep-water coral community in the Gulf of Mexico. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 20303–20308.

Whitfield CW, Behura SK, Berlocher SH*, et al.* (2006) Thrice out of africa: Ancient and recent expansions of the honey bee, *Apis mellifera*. *Science* **314**, 642-645.

Williams A, Schlacher TA, Rowden AA*, et al.* (2010) Seamount megabenthic assemblages fail to recover from trawling impacts. *Marine Ecology* **31**, 183-199.

CHAPTER 2

Genome-wide predictability of restriction sites across the eukaryotic tree of life

Santiago Herrera[1,2] Paula H. Reyes-Herrera[3], Timothy M. Shank [1]

[1] Biology Department, Woods Hole Oceanographic Institution, 266 Woods Hole Road, Woods Hole, MA 02543, USA

[2] MIT/WHOI Joint Program, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

[3] Facultad Ingeniería de Electrónica y Biomédica, Universidad Antonio Nariño, Carrera 3 Este # 47a -15, Bloque 4, Bogotá, Colombia.

Preprint available online: **Herrera S.**, P.H. Reyes-Herrera & T.M. Shank (2014) Genome-wide predictability of restriction sites across the eukaryotic tree of life. *bioRxiv,* doi: http://dx.doi.org/10.1101/007781. Git code repository: https://github.com/phrh/PredRAD

ABSTRACT

High-throughput sequencing of reduced representation libraries obtained through digestion with restriction enzymes – generally known as restriction site associated DNA sequencing (RAD-seq) – has become a common strategy to generate genome-wide genotypic and sequence data from eukaryotes. The choice of restriction enzyme is critical for the design of any RAD-seq study as it determines the number of genetic markers that can be obtained for a given taxon enabling a broad spectrum of applications, including marker discovery, population genomics, genomic mapping and phylogenetics. Here, we test the hypothesis that genome composition, in terms of GC content, and mono-, di- and trinucleotide composition, can be used to predict the number of restriction sites for a given combination of restriction enzyme and genome across the eukaryotic tree of life. Our analyses reveal that in most cases the trinucleotide genome composition model is the best predictor of the expected number of restriction sites in a eukaryotic genome, and the GC content and mononucleotide models the worst. We conclude that the predictability of restriction site frequencies in eukaryotic genomes needs to be treated on a case-specific

basis, whereby the phylogenetic position of the taxon of interest and the specific recognition sequence of the selected restriction enzyme are the chief foci among the most determinant factors. The software here developed, PredRAD (https://github.com/phrh/PredRAD), and the resulting databases constitute a valuable reference resource that will help guide the choice of restriction enzyme for any study using RAD-seq or related methods.

INTRODUCTION

The use of restriction enzymes to obtain reduced representation libraries from nuclear genomes, combined with the power of next-generation sequencing technologies, is rapidly becoming one of the most used strategies to generate genome-wide genotypic and sequence data in both model and non-model organisms (Baird *et al.* 2008; Andolfatto *et al.* 2011; Elshire *et al.* 2011; Peterson *et al.* 2012). The hundreds, thousands or tens of thousands of single nucleotide polymorphisms (SNPs) embedded in the resulting restriction site associated DNA (RAD) sequence tags (Miller *et al.* 2007; Baird *et al.* 2008) have a myriad of uses in biology ranging from genetic mapping (Wang *et al.* 2013; Weber *et al.* 2013) to population genomics (Hohenlohe *et al.* 2010; Andersen *et al.* 2012; White *et al.* 2013), phylogeography (Emerson *et al.* 2010; Reitzel *et al.* 2013), phylogenetics (Wagner *et al.* 2012; Eaton & Ree 2013), and SNP marker discovery (Scaglione *et al.* 2012; Toonen *et al.* 2013).

The choice of appropriate restriction enzyme(s) is critical for the effective design and application of RAD sequencing and a rapidly growing number of related methods such as genotyping-by-sequencing (Elshire *et al.* 2011), multiplexed shotgun genotyping (Andolfatto *et al.* 2011), double digest RAD-seq (Peterson *et al.* 2012), and ezRAD (Toonen *et al.* 2013). This choice determines the number of RAD markers that can be obtained, the amount of sequencing needed for a desired coverage level, the number of samples that can be multiplexed, the monetary cost, and ultimately the success of a project. It has been widely suggested that the number of restriction sites in a genome, for a given enzyme, can be roughly predicted using simple probability, if one has an estimate of the genome size and guanine-cytosine (GC) composition (Baird *et al.* 2008; Davey *et al.* 2011). Both of these parameters can be approximated in non-model organisms through sequencing-independent techniques such as flow cytometry (Vinogradov 1994, 1998; Šmarda *et al.* 2011). However, preliminary evidence suggests that there can be significant departures from expectations for particular combinations of taxa and restriction enzymes (Davey & Blaxter 2011; Davey *et al.* 2011).

Type II restriction enzymes, which are endonucleases chiefly produced by prokaryotic microorganisms, cleave double stranded DNA (dsDNA) at specific unmethylated recognition sequences that are 4 to 8 base pairs long and typically palindromic. These enzymes are thought to play an important role as defense systems against foreign phage dsDNA during infection or as selfish parasitic elements, and therefore have been the center of an evolutionary 'arms race' (Rambach & Tiollais 1974; Karlin *et al.* 1992; Rocha *et al.* 2001). Type II restriction enzymes are not known in eukaryotes and are not used as virulence factors by bacteria to infect eukaryotic hosts. Therefore there are no *a priori* reasons to believe that recognition sites in eukaryotic genomes are subject to selective pressures, but rather they should be evolutionarily neutral. Eukaryotic genomes have heterogeneous compositions with characteristic signatures at the level of di- and trinucleotides that are largely independent of coding status or function (Karlin & Mrázek 1997; Karlin *et al.* 1998; Gentles & Karlin 2001). Thus, it is possible that genome composition at these levels has a large influence on the abundance of short sequence patterns such as recognition sequences of restriction enzymes.

Here, we test the hypothesis that genome composition can be used to predict the number of restriction sites for a given combination of restriction enzyme and taxon. For this we: i) performed systematic *in silico* genome-wide surveys of restriction sites for diverse type II restriction enzymes in 434 eukaryotic whole and draft genomes to determine their frequencies across taxa; ii) examined the composition of genomes at the level of di- and trinucleotides to determine patterns of compositional biases among taxa; iii) developed stochastic models based on GC content, and mono-, di- and trinucleotide compositions to predict the frequencies of restriction sites across taxa and diverse kinds of type II restriction enzymes; iv) evaluated the accuracy of the predictive models by comparing the *in silico* observed frequencies of restriction sites to the expected frequencies predicted by the models. The number of restriction sites in a genome is not the only factor that determines the number of RAD loci that can be recovered experimentally. The architecture of each genome, and in particular the number of repetitive elements and gene duplications, can significantly decrease the number of unambiguous loci obtained via alignment to a reference genome or *de novo* assembly. To quantify this contribution we assessed the proportion of restriction-site associated DNA tags that can potentially be recovered unambiguously after empirical sequencing. The software here developed, PredRAD (https://github.com/phrh/PredRAD), and the resulting databases constitute a reference resource that will help guide the choice of restriction enzyme for any study using RAD-related methods.

RESULTS

Observed frequencies of restriction sites

To explore restriction site frequencies across the Eukaryotic tree of life we surveyed recognition sequences for 18 commonly used palindromic type II restriction enzymes in 434 whole and draft genomes. Observed frequencies of restriction sites were highly variable among broad taxonomic groups for the set of restriction enzymes here examined (Table 1) – except for FatI – with clear clustering patterns determined by phylogeny (Fig. 1). For example, with NgoMIV we observed 45.8 restriction sites per megabase (RS/Mb) ± 24.6 (mean ± SD) in core eudicot plants, compared to 277.4 ± 131.3 RS/Mb in commelinid plants (monocots). Among closely-related species the frequency patterns were similar and variability generally small. Observed frequencies of RS/Mb were inversely proportional to the length of the recognition sequence, with orders of magnitude differences among the 4-, 6-, and 8- cutters when compared within the same species. For example, in the starlet anemone *Nematostella vectensis* there were 3917.6, 167.6, and 6.9 RS/Mb for the 4-cutter FatI, 6-cutter PstI and 8-cutter SbfI, respectively. In contrast, nucleotide composition of the recognition sequence did not show a clear correlation with the observed frequency of restriction sites. For example, 83.6 RS/Mb ± 25.1 were observed in Neopterigii vertebrates for KpnI (GGTACC) and 622.6 RS/Mb ±119.1 were observed for PstI (CTGCAG), both recognition sequences with a GC content of 66.7%.

Dinucleotide compositional biases

Dinucleotide odds ratios ($\bar{\rho}^*_{XY}$) (Burge *et al.* 1992), a measurement of relative dinucleotide abundances given observed component frequencies used to explore genomic compositional biases, revealed significant compositional biases for all possible dinucleotides (Fig. 2). Both dinucleotides and trinucleotides are considered significantly underrepresented if the odds ratio is ≤ 0.78, significantly overrepresented if ≥ 1.23, and equal to expectation if =1 (Karlin *et al.* 1998). The dinucleotide compositional biases were highly variable among broad taxonomic groups (e.g., core eudicot plants) but generally similar within. Two dinucleotide complementary pairs, CG/GC and AT/TA, had highly dissimilar relative frequencies between the members of each pair. The largest biases were for CG, being significantly underrepresented in groups like core eudicot plants ($\bar{\rho}^*_{XY}$=0.68 ± 0.11), gnathostomate vertebrates ($\bar{\rho}^*_{XY}$=0.32 ± 0.12), the Pucciniales rust fungi ($\bar{\rho}^*_{XY}$=0.66 ± 0.08), gastropod mollusks ($\bar{\rho}^*_{XY}$=0.68, SD=0.01), the Trebouxiophyceae green algae ($\bar{\rho}^*_{XY}$=0.61 ± 0.19) and the Saccharomycetales yeast ($\bar{\rho}^*_{XY}$=0.78 ± 0.17). CG was significantly overrepresented in groups like the Apocrita insects

($\bar{\rho}^*_{XY}$=1.59 ± 0.18). The complementary dinucleotide GC was not particularly underrepresented in any broad taxonomic group, but tended towards overrepresentation in ecdysozoan invertebrates ($\bar{\rho}^*_{XY}$=1.24 ± 0.12), being significant in several arthropod and nematode species. Other taxa that showed significant overrepresentation of GC dinucleotides included the Trebouxiophyceae ($\bar{\rho}^*_{XY}$=1.39 ± 0.04) and microsporidia fungi ($\bar{\rho}^*_{XY}$=1.28 ± 0.17). Relative abundances of the dinucleotide AT were within expectations for all eukaryotes, except for the fungus *Sporobolomyces roseus* (ρ^*_{XY}=0.78). Contrastingly, the TA dinucleotide tended towards underrepresentation throughout the eukaryotes ($\bar{\rho}^*_{XY}$=0.8 ± 0.13), except in a few hypocreomycetid fungal species, for which it was significantly underrepresented. The TA dinucleotide was significantly underrepresented in trypanosomatids ($\bar{\rho}^*_{XY}$=0.59 ± 0.03), choanoflagellids ($\bar{\rho}^*_{XY}$=0.43 ± 0.09), chlorophytes ($\bar{\rho}^*_{XY}$=0.62 ± 0.15), stramenopiles ($\bar{\rho}^*_{XY}$=0.70 ± 0.07), and marginally underrepresented in most euteleostei fish ($\bar{\rho}^*_{XY}$=0.77 ± 0.04), archosaurs ($\bar{\rho}^*_{XY}$=0.76 ± 0.03) and the Basidiomycota ($\bar{\rho}^*_{XY}$=0.74 ± 0.09), among others.

The remaining dinucleotides had identical relative frequencies between the members of each complementary pair. The dinucleotide pair GG/CC was marginally underrepresented in most eukaryotes ($\bar{\rho}^*_{XY}$=0.88 ± 0.15). In the Sarcopterygii vertebrates ($\bar{\rho}^*_{XY}$=1.02 ± 0.06) and embryophyte plants ($\bar{\rho}^*_{XY}$=1.03 ± 0.07) GG/CC relative frequencies closely conformed to expectation, whereby GG/CC was significantly overrepresented in handful of isolated ecdysozoan, microsporidia and alveolate species, and significantly underrepresented in chlorophytes ($\bar{\rho}^*_{XY}$=0.72, SD=0.11), oomycetes ($\bar{\rho}^*_{XY}$=0.71 ± 0.05), and in several species of the Basidiomycota and the Dothideomycetes. Only the choanoflagellate *Salpingoeca* and the green alga *Asterochloris* presented a marginally significant bias for the dinucleotide pair AA/TT (ρ^*_{XY}=0.77 and 0.75 respectively). Similarly, *Salpingoeca* was the only taxon to show a significant bias for AC/GT (ρ^*_{XY}=1.42). Dinucleotide pair CA/TG was among the pairs with largest biases. Significant overrepresentation of CA/TG was found in several groups with large CG underrepresentation such as gnathostomates ($\bar{\rho}^*_{XY}$=1.31 ± 0.05), gastropods ($\bar{\rho}^*_{XY}$=1.29 ± 0.05), the Pucciniales ($\bar{\rho}^*_{XY}$=1.27 ± 0.02), the Trebouxiophyceae ($\bar{\rho}^*_{XY}$=1.62 ± 0.14), as well as several species of core eudicots and the Saccharomycetales. Other groups with significant CA/TG overrepresentation include onchocercid nematodes ($\bar{\rho}^*_{XY}$=1.26 ± 0.01), the Ustilaginomycotina fungi ($\bar{\rho}^*_{XY}$=1.28 ± 0.05), trypanosomatids ($\bar{\rho}^*_{XY}$=1.25 ± 0.04), and amoebozoans ($\bar{\rho}^*_{XY}$=1.33 ± 0.06). Overrepresentation biases for the AG/CT

dinucleotide pair were only present in amniotes ($\bar{\rho}^*_{XY}$=1.26 ± 0.02), the Sporidiobolales fungi ($\bar{\rho}^*_{XY}$=1.24 ± 0.01), and oxytrichid alveolates ($\bar{\rho}^*_{XY}$=1.24 ± 0.04), and other isolated species. Most of these taxa also had large CG underrepresentation. Lastly, most eukaryotes had GA/TC relative frequencies that conformed to expectations, except for few scattered species and small groups such as the Microbotryomycetes fungi ($\bar{\rho}^*_{XY}$=1.45 ± 0.13), the Mamiellales green algae ($\bar{\rho}^*_{XY}$=1.40 ± 0.08), and the Eimeriorina alveolates ($\bar{\rho}^*_{XY}$=1.26 ± 0.02).

Trinucleotide compositional biases

Trinucleotide odds ratios (γ^*_{XYZ}) (Burge *et al.* 1992) are another important measurement used to explore genomic compositional biases. Among the examined taxa, these ratios revealed compositional biases for most possible trinucleotides (Fig. 3). However, most of these biases were only significant in scattered individual species (Fig. 4). Among the trinucleotide pairs with significant underrepresentation, CTA/TAG and CGA/TCG showed the most definite broad taxonomic patterns. CTA/TAG was significantly underrepresented in most taxa, except for groups like commelinid plants (monocots) (γ^*_{XYZ}=0.87 ± 0.03), most core eudicots (γ^*_{XYZ}=0.81 ± 0.02), eleutherozoans (γ^*_{XYZ}=0.82 ± 0.01), molluscs (γ^*_{XYZ}=0.83 ± 0.01), and gnathostomates (γ^*_{XYZ}=0.82 ± 0.02) – exclusive of the chimaera *Callorhinchus milii*. Contrastingly, the trinucleotide CGA/TCG was only significantly underrepresented in most tetrapod vertebrates (γ^*_{XYZ}=0.82 ± 0.02) – exclusive of muroid rodents, bovid ruminants and the Afrotheria – a group containing aarvdvarks, hyraxes, and elephants.

The largest and more widespread overrepresentation biases were for the trinucleotide pair AAA/TTT, being significant in most eukaryotes, except for the majority of the Dikarya fungi (γ^*_{XYZ}=1.18 ± 0.07). The trinucleotide pairs TAA/TTA and AAT/ATT were significantly overrepresented in many metazoan taxa, particularly in the Neopterygii vertebrates (γ^*_{XYZ}=1.3 ± 0.05, and γ^*_{XYZ}=1.26 ± 0.05 respectively). AAG/CTT was significantly overrepresented in the Bacillariophyta diatoms (γ^*_{XYZ}=1.24 ± 0.03), oomycetes (γ^*_{XYZ}=1.28 ± 0.02), and the Saccharomycetales (γ^*_{XYZ}=1.26 ± 0.04). Lastly, CCA/TTG was significantly overrepresented in several tetrapod groups, including the Laurasiatheria – exclusive of the Chiroptera – (γ^*_{XYZ}=1.25 ± 0.02) and Hominoidea (γ^*_{XYZ}=1.23 ± 0.004).

Table 1. Restriction enzymes included in this study.

Core Sequence	Restriction Enzyme	Recognition Sequence	Recognition Sequence Length	GC Content of Recongition Sequence
GGCC				
	NotI	GCGGCCGC	8	100.0
CCGG				
	SgrAI	CRCCGGYG	8	87.5
	BsrFI	RCCGGY	6	83.3
	NgoMIV	GCCGGC	6	100.0
	AgeI	ACCGGT	6	66.7
	MspI	CCGG	4	100.0
TGCA				
	SbfI	CCTGCAGG	8	75.0
	PstI	CTGCAG	6	66.7
	NsiI	ATGCAT	6	33.3
AATT				
	ApoI	RAATTY	6	16.7
	EcoRI	GAATTC	6	33.3
	MluCI	AATT	4	0.0
TTAA				
	MseI	TTAA	4	0.0
CATG				
	NspI	RCATGY	6	50.0
	NcoI	CCATGG	6	66.7
	PciI	ACATGT	6	33.3
	FatI	CATG	4	50.0
GTAC				
	KpnI	GGTACC	6	66.7

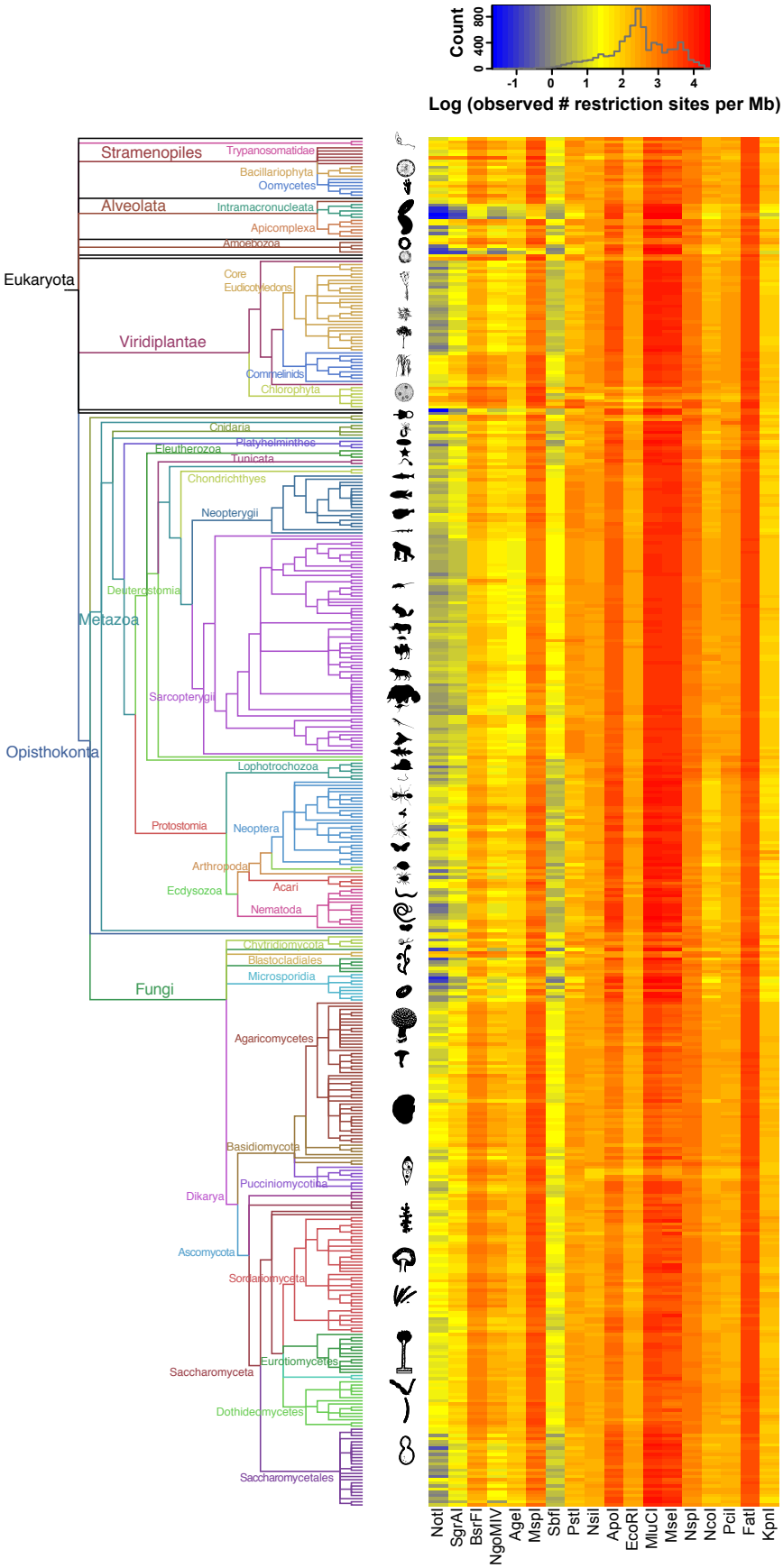

Count
0
400
800
-1
0
1
2
3
4
Log (observed # restriction sites per Mb)
Stramenopiles
Trypanosomatidae
Bacillariophyta
Oomycetes
Alveolata
Intramacronucleata
Apicomplexa
Amoebozoa
Eukaryota
Core
Eudicotyledons
Viridiplantae
Commelinids
Chlorophyta
Cnidaria
Platyhelminthes
Eleutherozoa
Tunicata
Chondrichthyes
Neopterygii
Deuterostomia
Metazoa
Sarcopterygii
Opisthokonta
Lophotrochozoa
Protostomia
Neoptera
Arthropoda
Ecdysozoa
Acari
Nematoda
Chytridiomycota
Blastocladiales
Microsporidia
Fungi
Agaricomycetes
Basidiomycota
Pucciniomycotina
Dikarya
Ascomycota
Sordariomycetes
Eurotiomycetes
Saccharomyceta
Dothideomycetes
Saccharomycetales
NotI
SgrAI
BsrFI
NgoMIV
AgeI
MspI
SbfI
PstI
NsiI
ApoI
EcoRI
MluCI
MseI
NspI
NcoI
PciI
FatI
KpnI

Figure 1. Observed restriction site frequencies. Left: phylogenetic tree of all eukaryotic taxa analyzed in this study. The tree is based on the NCBI taxonomy tree retrieved on May 16, 2013 using the iTOL tool http://itol.embl.de (Letunic & Bork 2011). Branch colors and labels indicate broad taxonomic groups. Organism silhouettes and cartoons were created by the authors or obtained from http://phylopic.org/. Right: heatmap of the observed frequency of restriction sites. Each row corresponds to a species from the tree on the left, and each column corresponds to a different restriction enzyme. Gray line in the color-scale box shows the distribution histogram of all values.

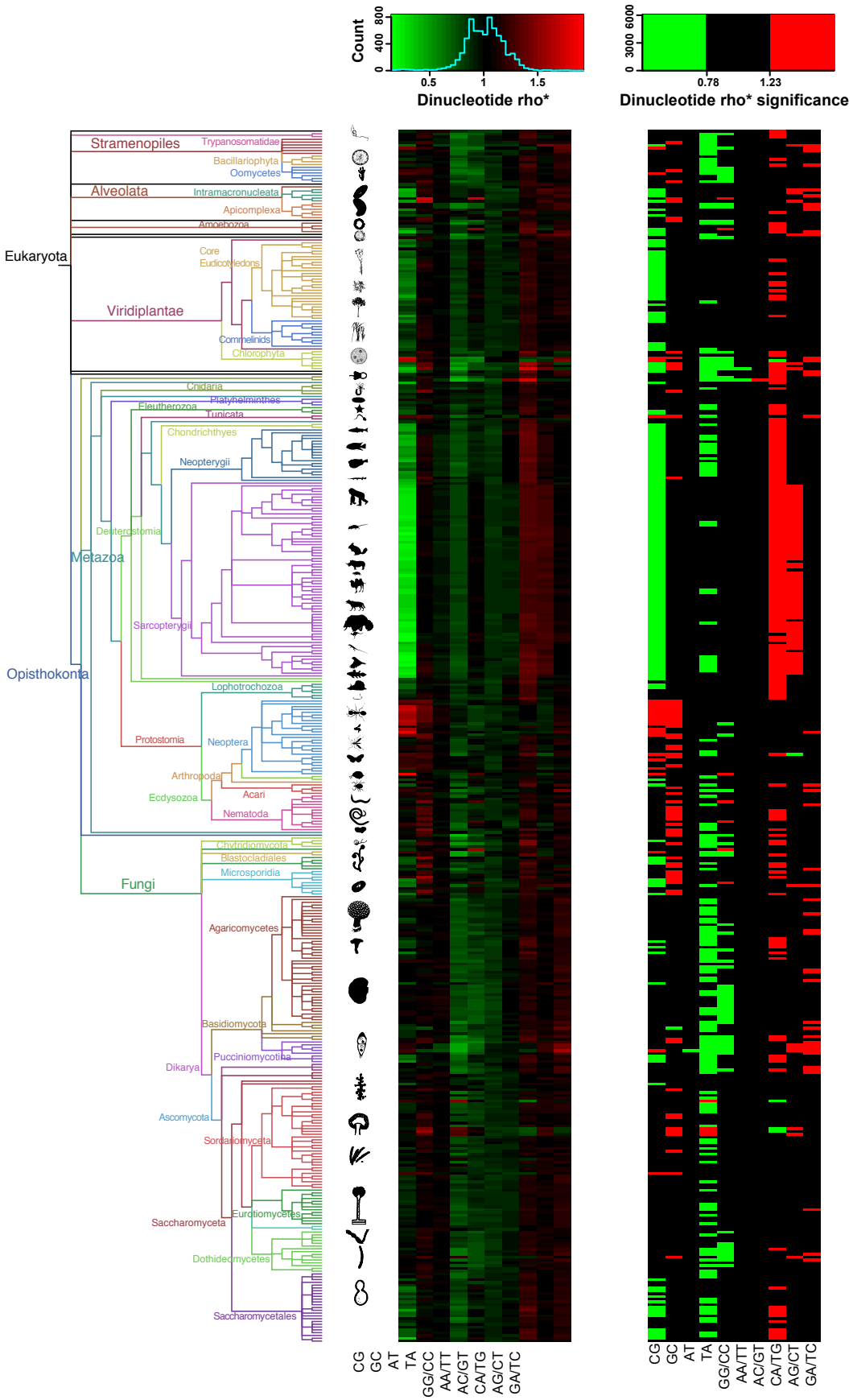

Count
0
400
800
0.5
1
1.5
Dinucleotide rho*
0
3000
6000
0.78
1.23
Dinucleotide rho* significance
Eukaryota
Stramenopiles
Trypanosomatidae
Bacillariophyta
Oomycetes
Alveolata
Intramacronucleata
Apicomplexa
Amoebozoa
Viridiplantae
Core Eudicotyledons
Commelinids
Chlorophyta
Opisthokonta
Metazoa
Cnidaria
Platyhelminthes
Eleutherozoa
Tunicata
Chondrichthyes
Neopterygii
Deuterostomia
Sarcopterygii
Lophotrochozoa
Protostomia
Neoptera
Arthropoda
Ecdysozoa
Acari
Nematoda
Fungi
Chytridiomycota
Blastocladiales
Microsporidia
Agaricomycetes
Basidiomycota
Pucciniomycotina
Dikarya
Ascomycota
Sordariomyceta
Eurotiomycetes
Saccharomyceta
Dothideomycetes
Saccharomycetales
CG
GC
AT
TA
GG/CC
AA/TT
AC/GT
CA/TG
AG/CT
GA/TC

Figure 2. Dinucleotide compositional biases and significances. Left: phylogenetic tree as in Fig. 1. Center: heatmap of the ρ^*_{XY} odds ratio values. Right: heatmap of the ρ^*_{XY} odds ratio significant values ρ^*_{XY}<0.78 and ρ^*_{XY}>1.23. Each row corresponds to a species from the tree on the left, and each column corresponds to a different dinucleotide. Green indicates underrepresentation and red indicates overrepresentation. Cyan line in the color-scale box shows the distribution histogram of all values.

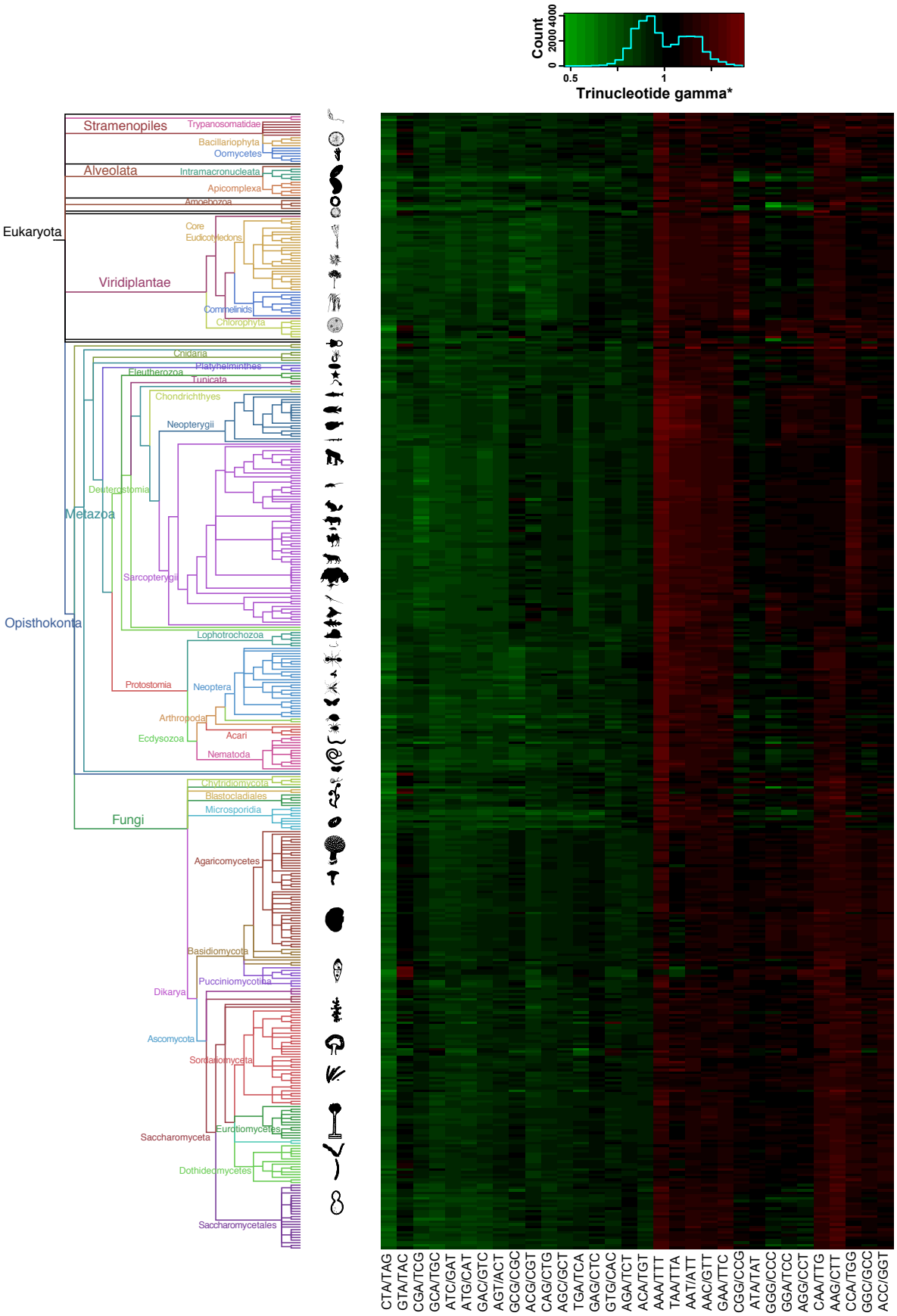

Count
0
2000
4000
0.5
1
Trinucleotide gamma*
Eukaryota
Stramenopiles
Trypanosomatidae
Bacillariophyta
Oomycetes
Alveolata
Intramacronucleata
Apicomplexa
Amoebozoa
Core Eudicotyledons
Viridiplantae
Commelinids
Chlorophyta
Cnidaria
Platyhelminthes
Eleutherozoa
Tunicata
Chondrichthyes
Neopterygii
Deuterostomia
Metazoa
Sarcopterygii
Opisthokonta
Lophotrochozoa
Protostomia
Neoptera
Arthropoda
Ecdysozoa
Acari
Nematoda
Chytridiomycota
Blastocladiales
Microsporidia
Fungi
Agaricomycetes
Basidiomycota
Pucciniomycotina
Dikarya
Ascomycota
Sordariomycetes
Eurotiomycetes
Saccharomyceta
Dothideomycetes
Saccharomycetales
CTA/TAG
GTA/TAC
CGA/TCG
GCA/TGC
ATC/GAT
ATG/CAT
GAC/GTC
AGT/ACT
GCG/CGC
ACG/CGT
CAG/CTG
AGC/GCT
TGA/TCA
GAG/CTC
GTG/CAC
AGA/TCT
ACA/TGT
AAA/TTT
TAA/TTA
AAT/ATT
AAC/GTT
GAA/TTC
CGG/CCG
ATA/TAT
GGG/CCC
GGA/TCC
AGG/CCT
CAA/TTG
AAG/CTT
CCA/TGG
GGC/GCC
ACC/GGT

Figure 3. Trinucleotide compositional biases. Left: phylogenetic tree as in Fig. 1. Right: heatmap of the γ^*_{XYZ} odds ratio values. Each row corresponds to a species from the tree on the left, and each column corresponds to a different trinucleotide. Green indicates underrepresentation and red indicates overrepresentation. Cyan line in the color-scale box shows the distribution histogram of all values.

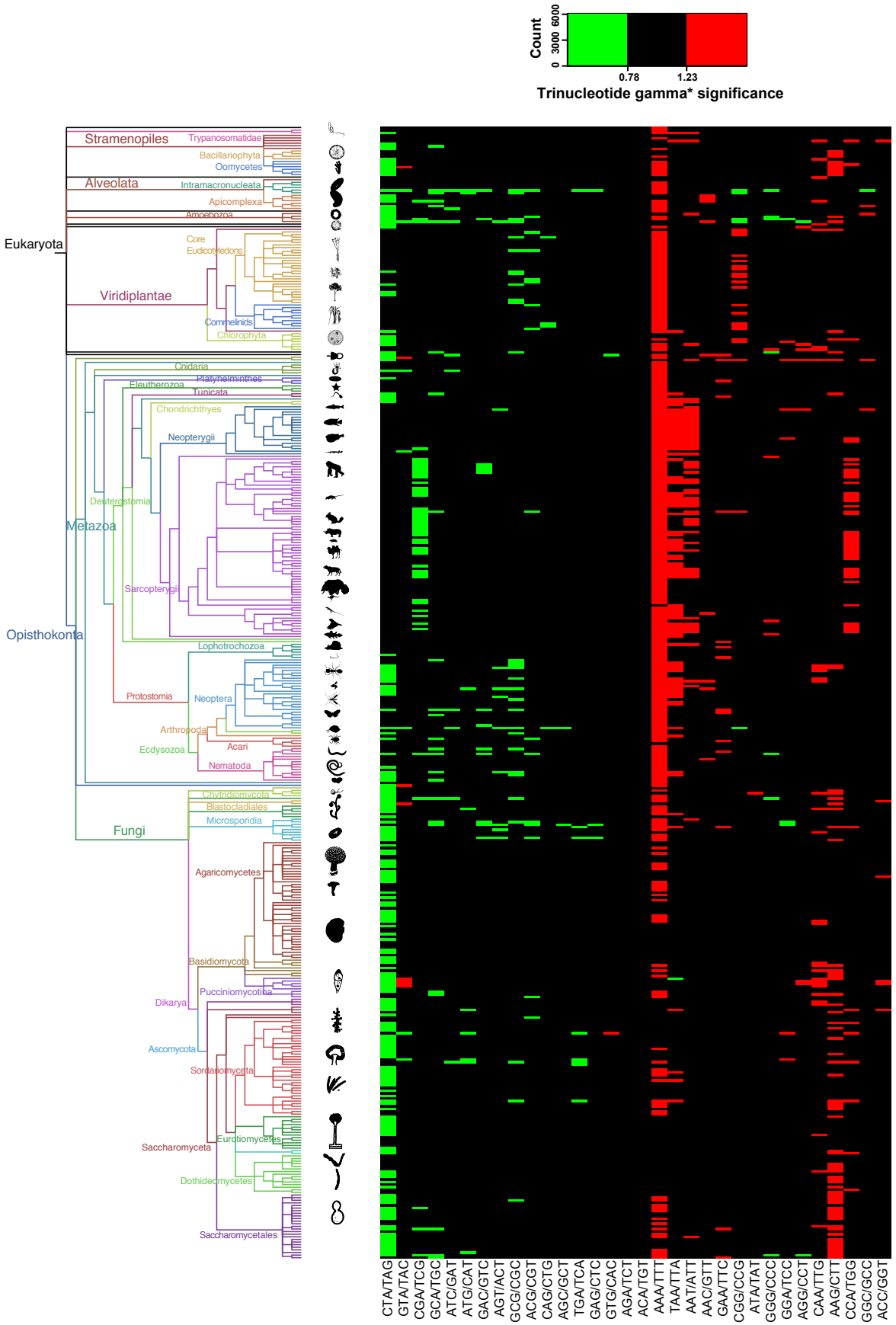

Count
6000
3000
0
0.78
1.23
Trinucleotide gamma* significance
Stramenopiles
Trypanosomatidae
Bacillariophyta
Oomycetes
Alveolata
Intramacronucleata
Apicomplexa
Amoebozoa
Eukaryota
Core Eudicotyledons
Viridiplantae
Commelinids
Chlorophyta
Cnidaria
Platyhelminthes
Eleutherozoa
Tunicata
Chondrichthyes
Neopterygii
Deuterostomia
Metazoa
Sarcopterygii
Opisthokonta
Lophotrochozoa
Protostomia
Neoptera
Arthropoda
Acari
Ecdysozoa
Nematoda
Chytridiomycota
Blastocladiales
Microsporidia
Fungi
Agaricomycetes
Basidiomycota
Pucciniomycotina
Dikarya
Ascomycota
Sordariomyceta
Eurotiomycetes
Saccharomyceta
Dothideomycetes
Saccharomycetales
CTA/TAG
GTA/TAC
CGA/TCG
GCA/TGC
ATC/GAT
ATG/CAT
GAC/GTC
AGT/ACT
GCG/CGC
ACG/CGT
CAG/CTG
AGC/GCT
TGA/TCA
GAG/CTC
GTG/CAC
AGA/TCT
ACA/TGT
AAA/TTT
TAA/TTA
AAT/ATT
AAC/GTT
GAA/TTC
CGG/CCG
ATA/TAT
GGG/CCC
GGA/TCC
AGG/CCT
CAA/TTG
AAG/CTT
CCA/TGG
GGC/GCC
ACC/GGT

Figure 4. Trinucleotide compositional biases significances. Left: phylogenetic tree as in Fig. 1. Right: heatmap of the γ^*_{XYZ} odds ratio significant values ρ^*_{XY}<0.78 and ρ^*_{XY}>1.23. Each row corresponds to a species from the tree on the left, and each column corresponds to a different trinucleotide. Green indicates underrepresentation and red indicates overrepresentation. Cyan line in the color-scale box shows the distribution histogram of all values.

Expected frequencies of restriction sites

To test the hypothesis that compositional heterogeneity in eukaryotic genomes can determine the frequency of restriction sites of each genome we developed probability models based on the GC content of each genome, as well as the mononucleotide, dinucleotide and trinucleotide compositions to predict the expected frequency of recognition sequences for each restriction enzyme. We evaluated the fit of each model using a similarity index (*SI),* defined as the quotient of the number of observed and expected restriction sites, minus one. A positive *SI* indicates that the number of observed restriction sites is greater than the expected, whereas a negative *SI* indicates a smaller number of observed sites than expected. If *SI* is equal to 0, then the number of observed sites is equal to the expectation. For example, a *SI* = 1 indicates that the number of observed restriction sites for a particular enzyme in a given genome is twice the number of expected sites predicted by a particular model. Trinucleotide composition models were in general a better predictor, in terms of their accuracy and precision, of the expected number of restriction sites than any of the other models (Fig. 5, Fig. 6). The mononucleotide and GC content models produced relatively poor predictions that were indistinguishable from one another (Fig. 5, Fig. 6). In a few cases the other models outperformed the trinucleotide model, e.g., EcoRI (Fig. 5, Fig. 6, Fig. 7). The fit of the predictions was highly variable among broad taxonomic groups but generally similar within, e.g., in Neopterigii vertebrates an average *SI* of 0.14 ± 0.19 for AgeI with the dinucleotide model, compared to -0.31 ± 0.19 in Sarcopterigii.

Recovery of RAD-tags after in silico sequencing

In most cases, the recovery of RAD-tags after *in silico* sequencing was notably high, with a median percentage of suppressed alignments to the reference genome assembly of only 3% (Fig. 8). There was no evident recovery bias by restriction enzyme, but rather bias was pronounced in a few individual species, likely indicating an enrichment of repetitive regions or duplications.

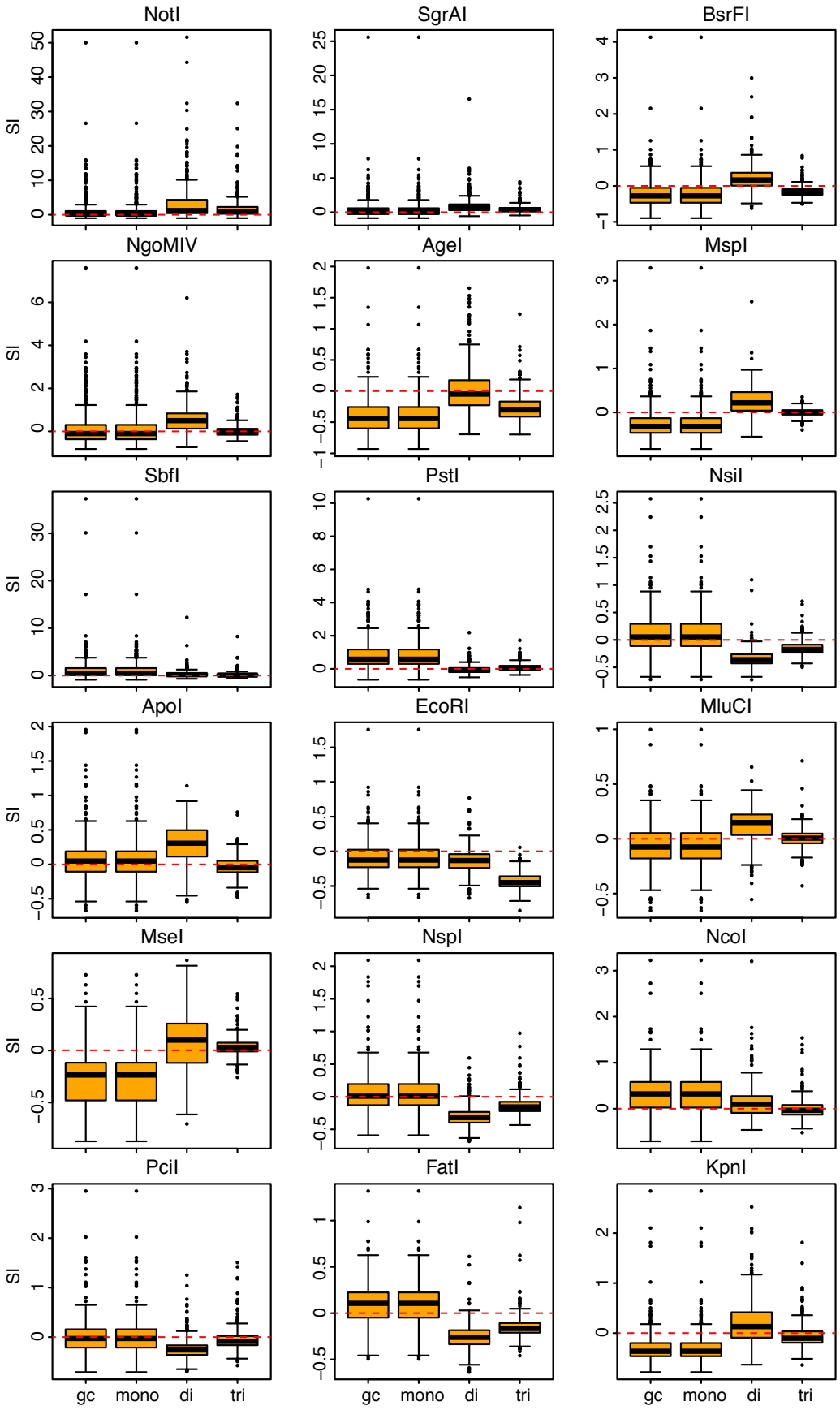

NotI
SgrAI
BsrFI
NgoMIV
AgeI
MspI
SbfI
PstI
NsiI
ApoI
EcoRI
MluCI
MseI
NspI
NcoI
PciI
FatI
KpnI
SI
gc
mono
di
tri

Figure 5. Overall fit of genome composition models per restriction enzyme. Vertical axes in the box and whisker plots indicate the values of the similarity index (*SI*) for each species per enzyme (see Methods section). Horizontal axes in the box and whisker plots indicate the genome composition model: GC content (gc), mononucleotide (mono), dinucleotide (di), and trinucleotide (tri). Horizontal edges of range boxes indicate the first and third quartiles of the *SI* values under each composition model. The thick horizontal black line represents the median. Whiskers indicate the value of 1.5 times the inter-quartile range from the first and third quartiles. Outliers are defined as SI values outside the whiskers range and are represented by dots. Outlier value of *Entamoeba histoyitica* for *NotI* was excluded. Red dotted lines indicate *SI*=0.

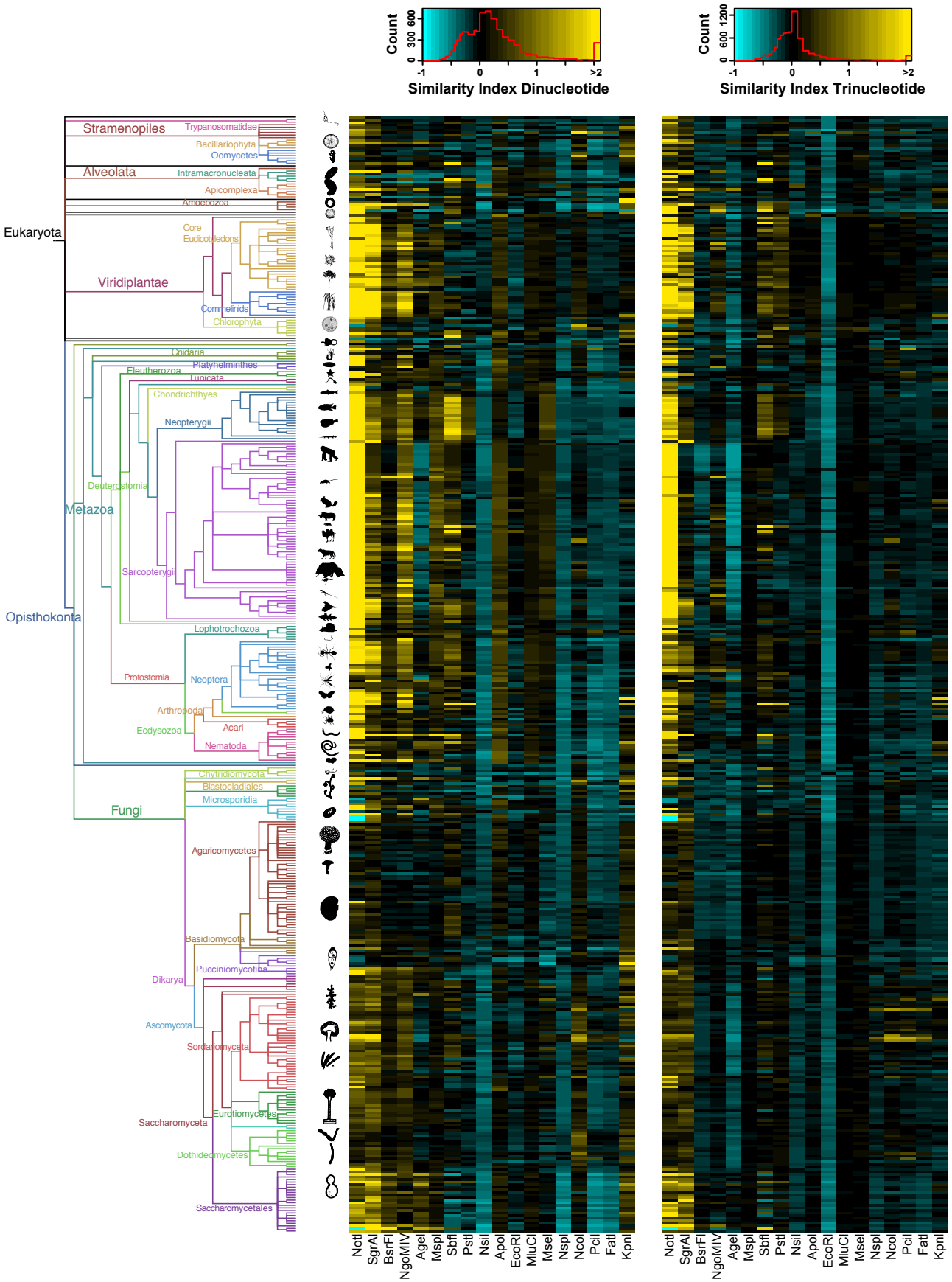

Count
300
600
0
-1
0
1
>2
Similarity Index Dinucleotide
Count
600
1200
0
-1
0
1
>2
Similarity Index Trinucleotide
Stramenopiles
Trypanosomatidae
Bacillariophyta
Oomycetes
Alveolata
Intramacronucleata
Apicomplexa
Amoebozoa
Eukaryota
Core
Eudicotyledons
Viridiplantae
Commelinids
Chlorophyta
Cnidaria
Platyhelminthes
Eleutherozoa
Tunicata
Chondrichthyes
Neopterygii
Deuterostomia
Metazoa
Sarcopterygii
Opisthokonta
Lophotrochozoa
Protostomia
Neoptera
Arthropoda
Ecdysozoa
Acari
Nematoda
Chytridiomycota
Blastocladiales
Microsporidia
Fungi
Agaricomycetes
Basidiomycota
Pucciniomycotina
Dikarya
Ascomycota
Sordariomycetes
Eurotiomycetes
Saccharomyceta
Dothideomycetes
Saccharomycetales
NotI
SgrAI
BsrFI
NgoMIV
AgeI
MspI
SbfI
PstI
NsiI
ApoI
EcoRI
MluCI
MseI
NspI
NcoI
PciI
FatI
KpnI

Figure 6. Similarity indexes for dinucleotide and trinucleotide genome composition models. Left: phylogenetic tree as in Fig. 1. Center: heatmap of the similarity indexes for the dinucleotide model Right: heatmap of the similarity indexes for the trinucleotide model**.** Each row corresponds to a species from the tree on the left, and each column corresponds to a different restriction enzyme. Cyan indicates $SI < 0$ and yellow indicates $SI > 0$. Red line in the color-scale box shows the distribution histogram of all values.

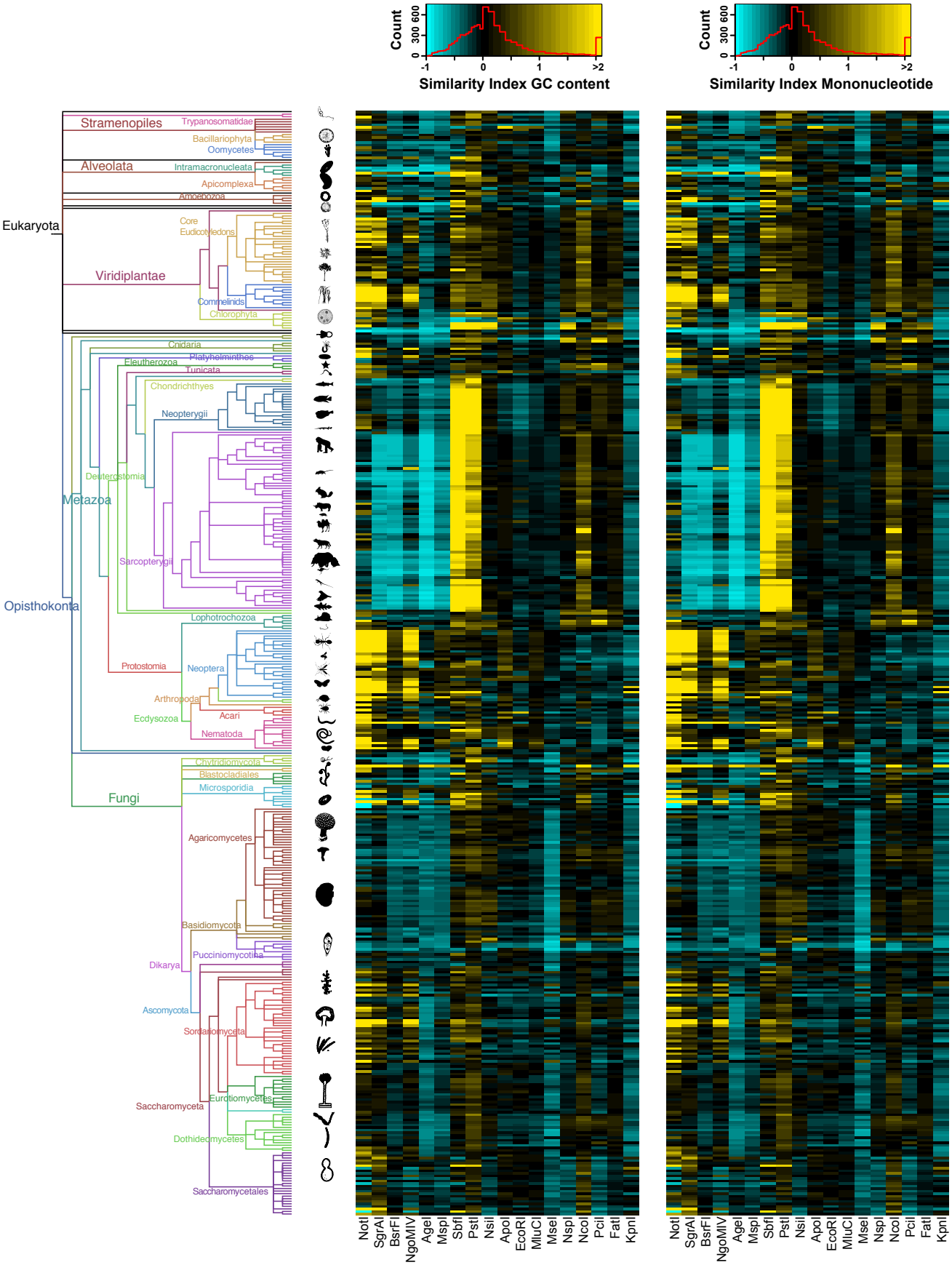

Count
0
300
600
-1
0
1
>2
Similarity Index GC content
Similarity Index Mononucleotide
Eukaryota
Stramenopiles
Trypanosomatidae
Bacillariophyta
Oomycetes
Alveolata
Intramacronucleata
Apicomplexa
Amoebozoa
Core Eudicotyledons
Viridiplantae
Commelinids
Chlorophyta
Cnidaria
Platyhelminthes
Eleutherozoa
Tunicata
Chondrichthyes
Neopterygii
Deuterostomia
Metazoa
Sarcopterygii
Opisthokonta
Lophotrochozoa
Protostomia
Neoptera
Arthropoda
Ecdysozoa
Acari
Nematoda
Chytridiomycota
Blastocladiales
Microsporidia
Fungi
Agaricomycetes
Basidiomycota
Pucciniomycotina
Dikarya
Ascomycota
Sordariomyceta
Eurotiomycetes
Saccharomyceta
Dothideomycetes
Saccharomycetales
NotI
SgrAI
BsrFI
NgoMIV
AgeI
MspI
SbfI
PstI
NsiI
ApoI
EcoRI
MluCI
MseI
NspI
NcoI
PciI
FatI
KpnI

Figure 7. Similarity indexes for GC content and mononucleotide genome composition models. Left: phylogenetic tree as in Fig. 1. Center: heatmap of the similarity indexes for the GC content model Right: heatmap of the similarity indexes for the mononucleotide model**.** Each row corresponds to a species from the tree on the left, and each column corresponds to a different restriction enzyme. Cyan indicates $SI < 0$ and yellow indicates $SI > 0$. Red line in the color-scale box shows the distribution histogram of all values.

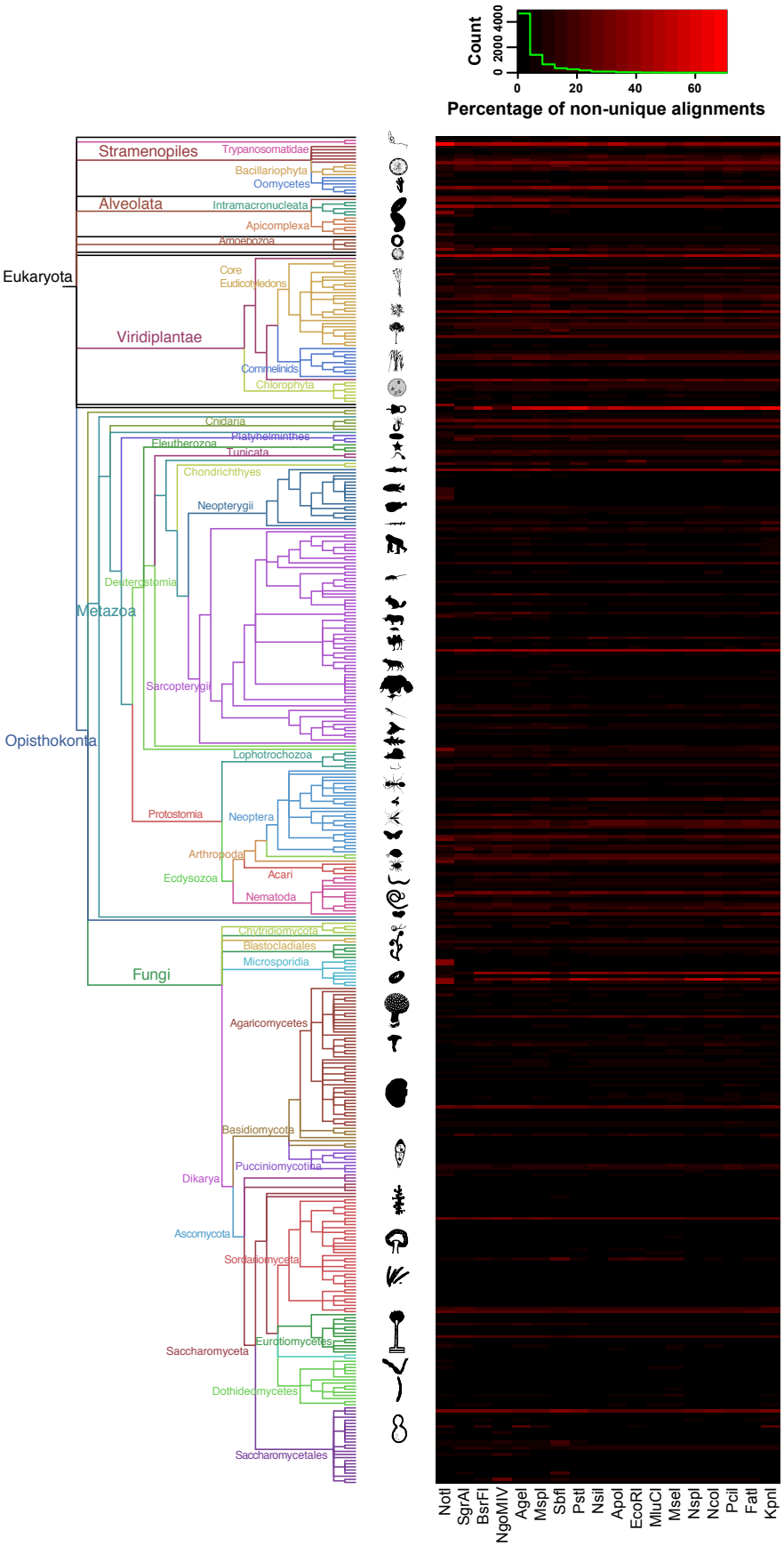

Count
0
2000
4000
0
20
40
60
Percentage of non-unique alignments
Stramenopiles
Trypanosomatidae
Bacillariophyta
Oomycetes
Alveolata
Intramacronucleata
Apicomplexa
Amoebozoa
Eukaryota
Core Eudicotyledons
Viridiplantae
Commelinids
Chlorophyta
Cnidaria
Platyhelminthes
Eleutherozoa
Tunicata
Chondrichthyes
Neopterygii
Deuterostomia
Metazoa
Sarcopterygii
Opisthokonta
Lophotrochozoa
Protostomia
Neoptera
Arthropoda
Acari
Ecdysozoa
Nematoda
Chytridiomycota
Blastocladiales
Microsporidia
Fungi
Agaricomycetes
Basidiomycota
Pucciniomycotina
Dikarya
Ascomycota
Sordariomyceta
Eurotiomycetes
Saccharomyceta
Dothideomycetes
Saccharomycetales
NotI
SgrAI
BsrFI
NgoMIV
AgeI
MspI
SbfI
PstI
NsiI
ApoI
EcoRI
MluCI
MseI
NspI
NcoI
PciI
FatI
KpnI

Figure 8. Recovery of RAD-tags after *in silico* genome digestion and sequencing. Left: phylogenetic tree as in Fig. 1. Right: heatmap of the percentage of RAD-tags that produced more than one unique alignment to their reference genome**.** Each row corresponds to a species from the tree on the left, and each column corresponds to a different restriction enzyme. Green line in the color-scale box shows the distribution histogram of all values.

DISCUSSION

Genome-wide surveys of restriction sites

Observed cut frequencies for a given restriction enzyme are strikingly variable among broad eukaryotic taxonomic groups, but similar among closely-related species. This is consistent with the hypothesis that the abundance of restriction sites is largely determined by phylogenetic relatedness. This pattern is most evident in groups that have a larger taxonomic representation, such as mammals. As more genome assemblies become available the pattern resolution will become clearer in many other underrepresented taxonomic groups, and through the use of comparative methods in a robust phylogenetic framework, it will be possible to establish taxon-specific divergence thresholds diagnostic of significant evolutionary changes in genome architecture.

As expected, observed frequencies of restriction sites with shorter recognition sequences are generally higher than the observed frequencies with longer recognition sequences. However this pattern in not universal. There are several instances in which the frequency of restriction sites for a high-denomination cutter is higher than for a low-denomination cutter. For example, in primates the frequency of the 8-cutter SbfI (24.6 ± 1.7 RS/Mb) is significantly higher than the frequency of the 6-cutter AgeI (18.4 ± 1.4 RS/Mb). These deviations from expectation are indicative of enzyme-specific frequency biases for particular taxa, and, as illustrated in the results section, are not correlated with the base composition of recognition sequences.

Genomic compositional biases

Our analyses indicate that there are significant compositional biases for most dinucleotides and trinucleotides across the eukaryotes. Many of these biases are significant only within individual species scattered throughout the eukaryotic tree of life. However, there are several particular dinuclotides and trinucleotides that show significant biases across the eukaryotic tree of life. Our observation that these biases are highly variable among broad taxonomic groups but generally similar within is congruent with findings from previous studies (e.g., Gentles & Karlin 2001). The most obvious biases across taxa are observed in the gnatostomate vertebrates; however, this is most likely due to rampant undersampling in

most other groups of eukaryotes (vertebrate genome assemblies represent 21% of all the taxa in this study).

The dinucleotides CG, GC, TA, and CA/TG show the most conspicuous bias patterns across the eukaryotic tree of life. Biases in most of these dinucleotides are likely linked to important biological processes. Notably the underrepresented dinucleotide CG is a widely known target for methylation related to transcriptional regulation (Bird 1980) and retrotransposon inactivation (Yoder *et al.* 1997) in vertebrates and eudicots. The corresponding overrepresentation of AG/CT fits the classic model of "methylation-deamination-mutation" by which a methylated cytosine in the CG pair tends to deaminate when unpaired and mutate into a thymidine with a corresponding CA complement. Interestingly CG and GC dinucleotides are significantly overrepresented in several groups of apocritic insects, as well as in some fungi and single-cell eukaryotes. CG is not a primary target for methylation in *Drosophila* (Lyko *et al.* 2000), instead CT, and in lesser degree CA and CC, are methylated in higher proportion. None of these dinucleotide pairs is significantly underrepresented in apocritic insects. The widespread TA underrepresentation has been traditionally attributed to stop codon biases, thermodynamic instability and susceptibility of UA to cleavage by RNAses in RNA transcripts (Beutler *et al.* 1989).

The trinucleotides CTA/TAG, AAA/TTT, TAA/TTA, CCA/TGG show the most conspicuous bias patterns across the eukaryotic tree of life. The biases in CTA/TAG have been widely attributed to the stop codon nature of UAG. However, the trinucleotides corresponding to the other stop codons (Burge *et al.* 1992), UAA and UGA, are overrepresented or not biased across eukaryotes. The reasons behind other cases of trinucleotide biases are less understood.

Predictability of restriction site frequencies

Our analyses indicate that in most cases the trinucleotide genome composition model is the best predictor, and the GC content and mononucleotide models are the worst predictors of the expected number of restriction sites in a eukaryotic genome. It is possible that the greater number of parameters in the trinucleotide model (64, compared to 16, 4 and 2 of the dinucleotide, mononucleotide and GC content model, respectively) is the cause of the better fit. However, this trend is not universal. As illustrated in the results section, the other models in a few cases outperformed the trinucleotide composition model. Neither the GC content nor length of the recognition sequence can confidently explain the observed discrepancies. It is not surprising that fit of the predictions made by the models is highly variable across taxonomic groups, given the high variability observed in the frequencies of restriction sites and genetic compositions across the eukaryotic tree of life. We conclude that the predictability of restriction site

frequencies in eukaryotic genomes needs to be treated on a case-specific basis, whereby the phylogenetic position of the taxon of interest and the specific recognition sequence of the selected restriction enzyme are the chief foci among the most determinant factors.

Implications for RAD-seq and related methodologies

For the design of a study using RAD-seq, or a related methodology, there are two fundamental questions that researchers face: i) what is the best restriction enzyme to use to obtain a desired number of RAD tags in the organism of interest? and ii) how many markers can be obtained with a particular enzyme in the organism of interest? The results from this study coupled with the developed software pipeline PredRAD, will allow any researcher to obtain an approximate answer to these questions.

In a best-case scenario for the practical design of a study using RAD-seq, or a related methodology, the species of interest is already included in the database presented here. In this case, the best proxy for the estimated number of RAD tags that could be obtained empirically would be twice the number of *in silico* observed restriction sites for each restriction enzyme (each restriction site is expected to produce two RAD tags, one in each direction from the restriction site) minus the number of suppressed read alignments to the reference genome assembly. For example, the predicted number of RAD tags for SbfI in starlet anemone *Nematostella vectensis* is 3,370, being highly similar to the range of RAD tags (2,300-2,800) obtained empirically by Reitzel *et al.* (2013). For library preparation protocols in which a fragment size selection step is done without a prior shearing step, e.g., ddRAD (Peterson *et al.* 2012) and ezRAD (Toonen *et al.* 2013), the size.select function of the software package SimRAD (Lepais & Weir 2014) constitutes a valuable complementary study-design tool. If a new genome assembly becomes available for the target species and/or the researcher wishes to evaluate an additional restriction enzyme, PredRAD can be re-executed with these data to quantify the number of restriction sites and the recovery potential, as well as to estimate the probability of the new recognition sequence based on genome composition models.

In the scenario that the genome sequence of the species of interest is not available, the best alternative is to look at the closest relative with a genome assembly. A range of approximate values for the number of RAD tags can be obtained from i) the number of *in silico* observed restriction sites in the closely related species; ii) the frequency of restriction sites in the closely related species, and the genome size of the species of interest; and iii) the probability of the recognition sequence for the enzyme(s) based on the best-fit genome composition model (*SI* closest to 0) from the closely related species and the genome size of the target species. The genome size of the species of interest can be estimated through sequencing-independent techniques such as flow cytometry (Vinogradov 1994, 1998; Šmarda *et al.* 2011). For

example, the predicted range in the number of RAD tags that could be obtained using SbfI in thoracican barnacles, a group for which no genome sequence assemblies are available, is 10,000-30,000, based on the observed frequency of the SbfI recognition sequence and its probability using a trinucleotide composition model in the genome of the crustacean *Daphnia pulex* (ranges of genome size for barnacles were obtained from the Animal Genome Size Database, http://ww.genomesize.com). Herrera *et al.* (Chapter 3) empirically obtained *ca.* 18,000 RAD tags for several species in this group. The possibility that the frequency of restriction sites and genome composition can be accurately estimated from alternative datasets such as a transcriptome is worth evaluating during restriction-enzyme selection for taxa without sequenced genomes.

Additional factors that can influence the number of RAD tag markers that can be obtained experimentally, and need to be considered during study design and data analysis steps, include: genome differences among individuals, level of heterozygosity, the amount of methylation in the genome, the number of repetitive regions and gene duplicates present in the target genome, the sensitivity of a particular restriction enzyme to methylation, the efficiency of the enzymatic digestion, the quality of library preparation and sequencing, the amount of sequencing, sequencing and library preparation biases, and the parameters used to clean, cluster and analyze the data, among others (see Davey *et al.* (2013), (Catchen *et al.* 2013), DaCosta and Sorenson (2014), and Mastretta-Yanes *et al.* (2014) for further discussion).

CONCLUSIONS

In this study we tested the hypothesis that genome composition can be used to predict the number of restriction sites for a given combination of restriction enzyme and genome. Our analyses reveal that in most cases the trinucleotide genome composition model is the best predictor, and the GC content and mononucleotide models are the worst predictors of the expected number of restriction sites in a eukaryotic genome. We conclude that the predictability of restriction site frequencies in eukaryotic genomes needs to be treated on a case-specific basis, whereby the phylogenetic position of the taxon of interest and the specific recognition sequence of the selected restriction enzyme are the chief foci among the most determinant factors. The results from this study and the software developed from it will help guide the design of any study using RAD sequencing and related methods. As more genome assemblies become available in underrepresented taxonomic groups the patterns of compositional biases and restriction site frequencies across the eukaryotic tree of life will become clearer and will improve our understanding of genome evolution.

METHODS

Observed frequencies of restriction sites

Assemblies from eukaryotic whole genome shotgun (WGS) sequencing projects available as of December 2012 were retrieved primarily from the U.S. National Center for Biotechnology Information (NCBI) WGS database (Additional file 1). Only one species per genus was included. Of the 434 genome assemblies included in this study, 42% corresponded to fungi, 21% to vertebrates, 16% invertebrates, and 9% plants. Only unambiguous nucleotide calls were taken into account. Genome sequence sizes were measured as the number of unambiguous nucleotides in the assembly. A set of 18 commonly used palindromic type II restriction enzymes with variable nucleotide compositions was screened in each of the genome assemblies (Table 1). The number of restriction sites present in each genome was obtained by counting the number of unambiguous matches for each recognition sequence pattern. Under optimal experimental conditions each restriction site should produce two RAD tags, one in each direction from the restriction site. Therefore, we define the number of observed RAD tags in each genome assembly as twice the number of recognition sequence pattern matches.

Expected frequencies of restriction sites

To test the hypothesis that compositional heterogeneity in eukaryotic genomes can determine the frequency of restriction sites of each genome, we characterized the GC content, as well as the mononucleotide, dinucleotide and trinucleotide compositions of each genome and developed probability models to predict the expected frequency of recognition sequences for each restriction enzyme. GC content was calculated as the proportion of unambiguous nucleotides in the assembly that are either guanine or cytosine, assuming that the frequency of guanine is equal to the frequency of cytosine. Mononucleotide composition was determined as the frequency of each one of the four nucleotides. Dinucleotide and trinucleotide compositions were determined as the frequency of each one of the 16 or 64 possible nucleotide combinations, respectively. The odds ratios proposed by Burge *et al.* (1992) were used to estimate compositional biases of dinucleotides (1) and trinucleotides (2) across genomes.

(1)

$$\rho^*_{XY} = \frac{f^*_{XY}}{f^*_X f^*_Y}$$

(2)

$$\gamma^*_{XYZ} = \frac{f^*_{XYZ} f^*_X f^*_Y f^*_Z}{f^*_{XY} f^*_{YZ} f^*_{XNZ}}$$

Where f_X^* is the relative frequency of the mononucleotide *X*, f_{XY}^* is the relative frequency of the dinucleotide *XY*, and f_{XYZ}^* is the relative frequency of the trinucleotide *XYZ*. All frequencies take into account the antiparallel structure of double stranded DNA. *N* represents any mononucleotide.

Mononucleotide and GC content sequence models were used to estimate the probability of a particular recognition sequence (3) assuming that each nucleotide is independent of the others and of its position on the recognition sequence. The GC content model assumes that the relative frequencies of guanine and cytosine in the genome sequence are equal. This model has only two parameters, the GC and AT frequencies. In the mononucleotide model there are four parameters, one for each of the four possible nucleotides.

(3)

$$p(s) = \prod_{i=1,\ldots,n(s)} p(s_i)$$

Here, $p(s_i)$ is the probability of nucleotide s_i at the position i of the recognition sequence. In the GC content model $p(s_i)$ can take the values of $f_{G,C}$ or $f_{A,T}$. In the mononucleotide model $p(s_i)$ can take the values of f_A, f_G, f_C, or f_T. Where f_X is the frequency of a given mononucleotide.

Dinucleotide and trinucleotide sequence models were defined as first and second degree Markov chain transition probability models with 16 or 64 parameters, respectively (Karlin *et al.* 1992; Singh 2009). These models take into account the position of each nucleotide in the recognition sequence. Nucleotides along the recognition sequence are not independent from nucleotides in neighboring positions. The probability of a particular recognition sequence for these Markov chain models was calculated as:

(4)

$$p(s) = p(s_1) \prod_{i=2,\ldots,n(s)} p_c(s_i|s_{i-1}, \ldots, s_{i-n})$$

Where $p(s_1)$ is the probability at the first position on the recognition sequence and p_c is the conditional probability of a subsequent nucleotide on the recognition sequence depending on the previous *n* nucleotides. In the dinucleotide sequence model $n = 1$ and in the trinucleotide sequence models $n = 2$.

Expectations versus observations

To assess the effectiveness of the predictive recognition sequence models, we compared the number of observed restriction sites in the genome assemblies with the expected number. The expected number of restriction sites in a given genome was calculated as the product of the probability of a recognition sequence multiplied by the genome sequence size. To quantify the departures from expectation, we define a similarity index (*SI)* as $SI = (O - E)/E$, where O and E are the observed and expected number of restriction sites, respectively. If $SI = 0$, then $E = O$. If $SI < 0$, then $E > O$, and *vice versa*.

Recovery of restriction-site associated DNA tags

To assess the proportion of restriction-site associated DNA tags that can potentially be recovered unambiguously after empirical sequencing, we performed *in silico* sequencing experiments for all genome assembly-restriction enzyme combinations. For each restriction site located in the genome assemblies, 100 base pairs up- and down-stream of the restriction site were extracted. This sequence read length is typical of sequencing experiments performed with current Hi-Seq platforms (Illumina Inc.). The resulting RAD tags were aligned back to their original genome assemblies using BOWTIE v0.12.7 (Langmead *et al.* 2009). Only reads that produced a unique best alignment were retained.

The analytical software pipeline here described and output database files are publicly available at https://github.com/phrh/PredRAD.

AUTHOR'S CONTRIBUTIONS

SH conceived and designed research. SH and PHR developed the software. SH analyzed the data. SH, TMS and HW contributed computing equipment. SH wrote the paper with comments from PHR and TMS.

ACKNOWLEDGEMENTS

This research was supported by the Office of Ocean Exploration and Research of the National Oceanic and Atmospheric Administration (NA09OAR4320129 to TS); the Division of Ocean Sciences of the National Science Foundation (OCE-1131620 to TS); the Astrobiology Science and Technology for Exploring Planets program of the National Aeronautics and Space Administration (NNX09AB76G to TS); and the Academic Programs Office (Ocean Ventures Fund to SH), the Ocean Exploration Institute

(Fellowship support to TMS) and the Ocean Life Institute of the Woods Hole Oceanographic Institution (internal grant to TMS and SH). We thank A. Reitzel, A. Tarrant, and C. Dunn for providing helpful discussions. A. Tarrant, A. Reitzel, E. Bors, A. Govindarajan, provided constructive comments on this manuscript.

REFERENCES

Andersen EC, Gerke JP, Shapiro JA*, et al.* (2012) Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nature Genetics* **44**, 285-290.

Andolfatto P, Davison D, Erezyilmaz D*, et al.* (2011) Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Research* **21**, 610-617.

Baird N, Etter P, Atwood T*, et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**, e3376.

Beutler E, Gelbart T, Han J, Koziol J, Beutler B (1989) Evolution of the genome and the genetic code: Selection at the dinucleo- tide level by methylation and polyribonucleotide cleavage. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 192-196.

Bird AP (1980) DNA methylation and the frequency of Cpg in animal DNA. *Nucleic Acids Research* **8**, 1499-1504.

Burge C, Campbell AM, Karlin S (1992) Over- and under-representation of short oligonucleotides in DNA sequences. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 1358-1362.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**, 3124-3140.

DaCosta JM, Sorenson MD (2014) Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. *PLoS One* **9**, e106713.

Davey JW, Blaxter ML (2011) RADSeq: Next-generation population genetics. *Briefings in Functional Genomics and Proteomics* **9**, 416-423.

Davey JW, Cezard T, Fuentes-Utrilla P*, et al.* (2013) Special features of RAD sequencing data: Implications for genotyping. *Molecular Ecology* **22**, 3151-3164.

Davey JW, Hohenlohe PA, Etter PD*, et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**, 499-510.

Eaton DAR, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology* **62**, 689-706.

Elshire RJ, Glaubitz JC, Sun Q*, et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) spproach for high diversity species. *PLoS One* **6**, e19379.

Emerson KJ, Merz CR, Catchen JM*, et al.* (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 16196-16200.

Gentles AJ, Karlin S (2001) Genome-scale compositional comparisons in eukaryotes. *Genome Research* **11**, 540-546.

Herrera S, Watanabe H, Shank TM (Chapter 3) Evolutionary and biogeographical patterns of barnacles from deep-sea hydrothermal vents.

Hohenlohe P, Bassham S, Etter P*, et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* **6**, e1000862.

Karlin S, Burge C, Campbell AM (1992) Statistical analyses of counts and distributions of restriction sites in DNA sequences. *Nucleic Acids Research* **20**, 1363-1370.

Karlin S, Campbell AM, Mrázek J (1998) Comparative DNA analysis across diverse genomes. *Annual Review of Genetics* **32**, 185-225.

Karlin S, Mrázek J (1997) Compositional differences within and between eukaryotic genomes. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 10227-10232.

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**, R25.

Lepais O, Weir JT (2014) SimRAD: an R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. *Molecular Ecology Resources* **14**, 1314-1321.

Letunic I, Bork P (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research* **39**, W475-W478.

Lyko F, Ramashoye BH, Jaenisch R (2000) DNA methylation in *Drosophila melanogaster*. *Nature* **408**, 538-540.

Mastretta-Yanes A, Arrigo N, Alvarez N*, et al.* (2014) Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources* **15**, 28–41.

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research* **17**, 240-248.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One* **7**, e37135.

Rambach A, Tiollais P (1974) Bacteriophage having EcoRI endonucleases sites only in the nonessential sites of the genome. *Proceedings of the National Academy of Sciences of the United States of America* **71**, 3927-3930.

Reitzel AM, Herrera S, Layden MJ, Martindale MQ, Shank TM (2013) Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular Ecology* **22**, 2953-2970.

Rocha EPC, Danchin A, Viari A (2001) Evolutionary role of restriction/modification systems as revealed by comparative genome analysis. *Genome Research* **11**, 946-958.

Scaglione D, Acquadro A, Portis E*, et al.* (2012) RAD tag sequencing as a source of SNP markers in *Cynara cardunculus* L. *BMC Genomics* **13**, 3.

Singh GB (2009) Stochastic Models for Biological Patterns. In: *Bioinformatics for Systems Biology* (ed. Krawetz S), pp. 151-162. Springer, New York.

Šmarda P, Bureš P, Šmerda J, Horová L (2011) Measurements of genomic GC content in plant genomes with flow cytometry: a test for reliability. *New Phytologist* **193**, 513-521.

Toonen RJ, Puritz JB, Forsman ZH*, et al.* (2013) ezRAD: A simplified method for genomic genotyping in non-model organisms. *PeerJ* **1**, e203.

Vinogradov A (1994) Measurement by flow cytometry of genomic AT/GC ratio and genome size. *Cytometry* **16**, 34-40.

Vinogradov A (1998) Genome size and GC-percent in vertebrates as determined by flow cytometry: The triangular relationship. *Cytometry* **31**, 100-109.

Wagner CE, Keller I, Wittwer S*, et al.* (2012) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology* **22**, 787-798.

Wang J, Wurm Y, Nipitwattanaphon M*, et al.* (2013) A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493**, 664–668.

Weber JN, Peterson BK, Hoekstra HE (2013) Discrete genetic modules are responsible for complex burrow evolution in *Peromyscus* mice. *Nature* **493**, 402-405.

White TA, Perkins SE, Heckel G, Searle JB (2013) Adaptive evolution during an ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in Ireland. *Molecular Ecology* **22**, 2971-2985.

Yoder JA, Walsh CP, Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics* **13**, 335-340.

SUPPORTING INFORMATION

Table S1. Genome assemblies included in this study. Note that web addresses to individual assembly files, and the assembly files themselves, were current as of December 2012 and may have changed.

ID	Species	Genome Assembly
Acacas	*Acanthamoeba castellanii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEYA01.fasta.gz
Acitak	*Aciculosporium take*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFQZ01.fasta.gz
Aciric	*Acidomyces richmondensis*	http://genome.jgi-psf.org/Aciri1_iso/download/Aciri1_iso_AssemblyScaffolds.fasta.gz
Acralc	*Acremonium alcalophilum*	http://genome.jgi-psf.org/Acral2/download/Acral2_AssemblyScaffolds.fasta.gz
Acrech	*Acromyrmex echinatior*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEVX01.fasta.gz
Acrdig	*Acropora digitifera*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=BACK01.fasta.gz
Acypis	*Acyrthosiphon pisum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABLF02.fasta.gz
Aedaeg	*Aedes aegypti*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAGE02.fasta.gz
Agabis	*Agaricus bisporus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEOK01.fasta.gz
Ailmel	*Ailuropoda melanoleuca*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACTA01.fasta.gz
Ajecap	*Ajellomyces capsulatus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAJI01.fasta.gz
Alamos	*Alatina moseri*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AHZO01.fasta.gz
Albcan	*Albugo candida*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CAJG01.fasta.gz
Allmis	*Alligator mississippiensis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AKHW01.fasta.gz
Allmac	*Allomyces macrogynus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACDU01.fasta.gz
Altarb	*Alternaria arborescens*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AIIC01.fasta.gz
Amamus	*Amanita muscaria*	http://genome.jgi-psf.org/Amamu1/download/Amamu1_AssemblyScaffolds.fasta.gz
Ampque	*Amphimedon queenslandica*	ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes/Eukaryotes/invertebrates/Amphimedon_queenslandica/v1.0/l
Annalg	*Anncaliia algerae*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CAIR01.fasta.gz
Anocar	*Anolis carolinensis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAWZ02.fasta.gz
Anogam	*Anopheles gambiae*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAAB01.fasta.gz
Apimon	*Apiospora montagnei*	http://genome.jgi-psf.org/Apimo1/download/Apimo1_AssemblyScaffolds.fasta.gz
Apimel	*Apis mellifera*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AADG06.fasta.gz
Aplcal	*Aplysia californica*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AASC02.fasta.gz
Aratha	*Arabidopsis thaliana*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFNA01.fasta.gz
Artoli	*Arthrobotrys oligospora*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADOT01.fasta.gz
Artben	*Arthroderma benhamiae*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABSU01.fasta.gz
Ascsuu	*Ascaris suum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AMPH01.fasta.gz
Ascrub	*Ascoidea rubescens*	http://genome.jgi-psf.org/Ascru1/download/Ascru1_AssemblyScaffolds.fasta.gz
Ascapi	*Ascosphaera apis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AARE01.fasta.gz
Aspfum	*Aspergillus fumigatus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAHF01.fasta.gz
Ast_sp	*Asterochloris sp*	http://genome.jgi-psf.org/Astpho1/download/Astpho1_genomic_scaffolds.fasta.gz
Attcep	*Atta cephalotes*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADTU01.fasta.gz
Aurlim	*Aurantiochytrium limacinum*	http://genome.jgi-psf.org/Aurli1/download/Aurli1_AssemblyScaffolds.fasta.gz
Aurpul	*Aureobasidium pullulans*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AMCU01.fasta.gz
Aurano	*Aureococcus anophagefferens*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACJI01.fasta.gz
Aurdel	*Auricularia delicata*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFVO01.fasta.gz
Babbov	*Babesia bovis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAXT01.fasta.gz
Babino	*Babjeviella inositovora*	http://genome.jgi-psf.org/Babin1/download/Babin1_AssemblyScaffolds.fasta.gz

Batden	*Batrachochytrium dendrobatidis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADAR01.fasta.gz
Baucom	*Baudoinia compniacensis*	http://genome.jgi-psf.org/Bauco1/download/Bauco1_AssemblyScaffolds.fasta.gz
Beabas	*Beauveria bassiana*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADAH01.fasta.gz
Betnan	*Betula nana*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CAOK01.fasta.gz
Bignat	*Bigelowiella natans*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADNK01.fasta.gz
Bjeadu	*Bjerkandera adusta*	http://genome.jgi-psf.org/Bjead1_1/download/Bjead1_1_AssemblyScaffolds.fasta.gz
Blahom	*Blastocystis hominis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CABX01.fasta.gz
Blugra	*Blumeria graminis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABSB02.fasta.gz
Bomter	*Bombus terrestris*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AELG01.fasta.gz
Bommor	*Bombyx mori*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=BABH01.fasta.gz
Bostau	*Bos taurus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAFC03.fasta.gz
Botbot	*Botryobasidium botryosum*	http://genome.jgi-psf.org/Botbo1/download/Botbo1_AssemblyScaffolds.fasta.gz
Botfuc	*Botryotinia fuckeliana*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ALOC01.fasta.gz
Bradis	*Brachypodium distachyon*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADDN01.fasta.gz
Braflo	*Branchiostoma floridae*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABEP02.fasta.gz
Brarap	*Brassica rapa*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AENI01.fasta.gz
Brumal	*Brugia malayi*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAQA01.fasta.gz
Burxyl	*Bursaphelenchus xylophilus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CADV01.fasta.gz
Caeele	*Caenorhabditis elegans*	ftp://ftp.wormbase.org/pub/wormbase/genomes/c_elegans/sequences/dna/c_elegans.WS201.dna.fa.gz
Cajcaj	*Cajanus cajan*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFSP01.fasta.gz
Caljac	*Callithrix jacchus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACFV01.fasta.gz
Calmil	*Callorhinchus milii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAVX01.fasta.gz
Camfer	*Camelus ferus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AGVR01.fasta.gz
Camflo	*Camponotus floridanus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEAB01.fasta.gz
Canalb	*Candida albicans*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAFO01.fasta.gz
Canlup	*Canis lupus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAEX03.fasta.gz
Cansat	*Cannabis sativa*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AGQN01.fasta.gz
Captel	*Capitella teleta*	ftp://ftp.jgi-psf.org/pub/JGI_data/Capitella/v1.0/Capitella_spl.fasta.gz
Capowc	*Capsaspora owczarzaki*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACFS01.fasta.gz
Carpap	*Carica papaya*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABIM01.fasta.gz
Catang	*Catenaria anguillulae*	http://genome.jgi-psf.org/Catan1/download/Catan1_AssemblyScaffolds.fasta.gz
Cavpor	*Cavia porcellus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAKN02.fasta.gz
Cersim	*Ceratotherium simum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AKZM01.fasta.gz
Cerzea	*Cercospora zeaemaydis*	http://genome.jgi-psf.org/Cerzm1/download/Cerzm1_AssemblyScaffolds.fasta.gz
Cersub	*Ceriporiopsis subvermispora*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEOV01.fasta.gz
Chathe	*Chaetomium thermophilum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADUW01.fasta.gz
Chilan	*Chinchilla lanigera*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AGCD01.fasta.gz
Chlrei	*Chlamydomonas reinhardtii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABCN01.fasta.gz
Chlvar	*Chlorella variabilis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADIC01.fasta.gz
Chohof	*Choloepus hoffmanni*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABVD01.fasta.gz
Chrpic	*Chrysemys picta bellii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AHGY01.fasta.gz
Chrasi	*Chrysochloris asiatica*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AMDV01.fasta.gz
Cioint	*Ciona intestinalis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AABS01.fasta.gz

Citlan	*Citrullus lanatus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AGCB01.fasta.gz
Citsin	*Citrus sinensis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AJPS01.fasta.gz
Clagra	*Cladonia grayi*	http://genome.jgi-psf.org/Clagr2/download/Clagr2_genomic_scaffolds.fasta.gz
Clasph	*Cladosporium sphaerospermum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AIIA01.fasta.gz
Clafus	*Claviceps fusiformis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFRA01.fasta.gz
Clalus	*Clavispora lusitaniae*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAFT01.fasta.gz
Closin	*Clonorchis sinensis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=BADR02.fasta.gz
Cocpos	*Coccidioides posadasii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACFW01.fasta.gz
Cocsub	*Coccomyxa subellipsoidea*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AGSI01.fasta.gz
Coclun	*Cochliobolus lunatus*	http://genome.jgi-psf.org/Coclu2/download/Coclu2_AssemblyScaffolds.fasta.gz
Coerev	*Coemansia reversa*	http://genome.jgi-psf.org/Coere1/download/Coere1_AssemblyScaffolds.fasta.gz
Colglo	*Colletotrichum gloeosporioides*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ANPB01.fasta.gz
Concri	*Condylura cristata*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AJFV01.fasta.gz
Concor	*Conidiobolus coronatus*	http://genome.jgi-psf.org/Conco1/download/Conco1_AssemblyScaffolds.fasta.gz
Conput	*Coniophora puteana*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEIT01.fasta.gz
Conapo	*Coniosporium apollinis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AJKL01.fasta.gz
Copcin	*Coprinopsis cinerea*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AACS02.fasta.gz
Cormil	*Cordyceps militaris*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEVU01.fasta.gz
Cragig	*Crassostrea gigas*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFTI01.fasta.gz
Crigri	*Cricetulus griseus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFTD01.fasta.gz
Croque	*Cronartium quercuum*	http://genome.jgi-psf.org/Croqu1/download/Croqu1_AssemblyScaffolds.fasta.gz
Crypa2	*Cryphonectria parasitica*	http://genome.jgi-psf.org/Crypa2/download/Cryphonectria_parasiticav2.nuclearAssembly.unmasked.gz
Crygat	*Cryptococcus gattii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAFP01.fasta.gz
Crypa1	*Cryptosporidium parvum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAEE01.fasta.gz
Cucsat	*Cucumis sativus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACHR01.fasta.gz
Culqui	*Culex quinquefasciatus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAAB01.fasta.gz
Cyamer	*Cyanidioschyzon merolae*	http://www.ebi.ac.uk/ena/data/view/AP006483.2,AP006492.2,AP006493.2,AP006494.2,AP006495.2,AP(
Cybjad	*Cyberlindnera jadinii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=BAEL01.fasta.gz
Dac_sp	*Dacryopinax sp*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEUS01.fasta.gz
Dalesc	*Daldinia eschscholzii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AIID01.fasta.gz
Danple	*Danaus plexippus*	ftp://ftp.ensemblgenomes.org/pub/metazoa/release-16/fasta/danaus_plexippus/dna/Danaus_plexippus.D
Danrer	*Danio rerio*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CABZ01.fasta.gz
Dappul	*Daphnia pulex*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACJG01.fasta.gz
Dasnov	*Dasypus novemcinctus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAGV03.fasta.gz
Daumad	*Daubentonia madagascariensis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AGTM01.fasta.gz
Debhan	*Debaryomyces hansenii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AHBE01.fasta.gz
Dekbru	*Dekkera bruxellensis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AHIQ01.fasta.gz
Diclab	*Dicentrarchus labrax*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CABK01.fasta.gz
Dicsqu	*Dichomitus squalens*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEID01.fasta.gz
Dicdis	*Dictyostelium discoideum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAFI02.fasta.gz
Didexi	*Didymella exigua*	http://genome.jgi-psf.org/Didex1/download/Didex1_AssemblyScaffolds.fasta.gz
Dipord	*Dipodomys ordii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABRO01.fasta.gz
Dromel	*Drosophila melanogaster*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AABU01.fasta.gz

Echtel	*Echinops telfairi*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAIY02.fasta.gz
Ectsil	*Ectocarpus siliculosus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CABU01.fasta.gz
Edhaed	*Edhazardia aedis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFBI02.fasta.gz
Eleedw	*Elephantulus edwardii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AMGZ01.fasta.gz
Emihux	*Emiliania huxleyi*	http://genome.jgi-psf.org/Emihu1/download/Emihu1_scaffolds.fasta.gz
Enccun	*Encephalitozoon cuniculi*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEWD01.fasta.gz
Enthis	*Entamoeba histolytica*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAFB02.fasta.gz
Entbie	*Enterocytozoon bieneusi*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABGB01.fasta.gz
Epityp	*Epichloe typhina*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AMDI01.fasta.gz
Eptfus	*Eptesicus fuscus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ALEH01.fasta.gz
Equcab	*Equus caballus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAWR02.fasta.gz
Erieur	*Erinaceus europaeus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AMDU01.fasta.gz
Erypis	*Erysiphe pisi*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CACM01.fasta.gz
Euccam	*Eucalyptus camaldulensis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=BADO01.fasta.gz
Eurher	*Eurotium herbariorum*	http://genome.jgi-psf.org/Eurhe1/download/Eurhe1_AssemblyScaffolds.fasta.gz
Eutpar	*Eutrema parvulum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFAN01.fasta.gz
Exoder	*Exophiala dermatitidis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFPA01.fasta.gz
Felcat	*Felis catus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AANG02.fasta.gz
Fibrad	*Fibroporia radiculosa*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CAGC01.fasta.gz
Ficalb	*Ficedula albicollis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AGTO01.fasta.gz
Fommed	*Fomitiporia mediterranea*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEJJ01.fasta.gz
Fompin	*Fomitopsis pinicola*	http://genome.jgi-psf.org/Fompi3/download/Fompi3_AssemblyScaffolds.fasta.gz
Fraves	*Fragaria vesca*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEMH01.fasta.gz
Fracyl	*Fragilariopsis cylindrus*	http://genome.jgi-psf.org/Fracy1/download/portalData/Fracy1_assembly_scaffolds.fasta.gz
Fusoxy	*Fusarium oxysporum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAXH01.fasta.gz
Gadmor	*Gadus morhua*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CAEA01.fasta.gz
Gaegra	*Gaeumannomyces graminis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADBI01.fasta.gz
Galmar	*Galerina marginata*	http://genome.jgi-psf.org/Galma1/download/Galma1_AssemblyScaffolds.fasta.gz
Galgal	*Gallus gallus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AADN03.fasta.gz
Ganluc	*Ganoderma lucidum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AGAX01.fasta.gz
Gasacu	*Gasterosteus aculeatus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AANH01.fasta.gz
Geodes	*Geomyces destructans*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEFC01.fasta.gz
Geofor	*Geospiza fortis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AKZB01.fasta.gz
Gialam	*Giardia lamblia*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AACB02.fasta.gz
Gibmon	*Gibberella moniliformis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAIM02.fasta.gz
Glaloz	*Glarea lozoyensis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AGUE01.fasta.gz
Glotra	*Gloeophyllum trabeum*	http://genome.jgi-psf.org/Glotr1_1/download/Glotr1_1_AssemblyScaffolds.fasta.gz
Glogra	*Glomerella graminicola*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACOD01.fasta.gz
Glymax	*Glycine max*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACUP01.fasta.gz
Gonpro	*Gonapodya prolifera*	http://genome.jgi-psf.org/Ganpr1/download/Ganpr1_AssemblyScaffolds.fasta.gz
Gorgor	*Gorilla gorilla*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CABD02.fasta.gz
Gosrai	*Gossypium raimondii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ALYE01.fasta.gz
Grocla	*Grosmannia clavigera*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACXQ02.fasta.gz

Guithe	*Guillardia theta*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEIE01.fasta.gz
Gymlux	*Gymnopus luxurians*	http://genome.jgi-psf.org/Gymlu1/download/Gymlu1_AssemblyScaffolds.fasta.gz
Hamtva	*Hamiltosporidium tvaerminnensis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACSZ01.fasta.gz
Hamham	*Hammondia hammondi*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AHJH01.fasta.gz
Hanval	*Hanseniaspora valbyensis*	http://genome.jgi-psf.org/Hanva1_1/download/Hanva1_1_AssemblyScaffolds.fasta.gz
Hapbur	*Haplochromis burtoni*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFNZ01.fasta.gz
Harsal	*Harpegnathos saltator*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEAC01.fasta.gz
Hebcyl	*Hebeloma cylindrosporum*	http://genome.jgi-psf.org/Hebcy2/download/Hebcy2_AssemblyScaffolds.fasta.gz
Helmel	*Heliconius melpomene*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CAEZ01.fasta.gz
Helrob	*Helobdella robusta*	ftp://ftp.jgi-psf.org/pub/JGI_data/Helobdella_robusta/v1.0/Helobdella_robusta.fasta.gz
Hetirr	*Heterobasidion irregulare*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEOJ01.fasta.gz
Hetgla	*Heterocephalus glaber*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AHKG01.fasta.gz
Hetgly	*Heterodera glycines*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABLA01.fasta.gz
Hetbac	*Heterorhabditis bacteriophora*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACKM01.fasta.gz
Homsap	*Homo sapiens*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AMYH01.fasta.gz
Hompol	*Homoloaphlyctis polyrhiza*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFSM01.fasta.gz
Horvul	*Hordeum vulgare*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CAJW01.fasta.gz
Hyaara	*Hyaloperonospora arabidopsidis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABWE02.fasta.gz
Hydpin	*Hydnomerulius pinastri*	http://genome.jgi-psf.org/Hydpi2/download/Hydpi2_AssemblyScaffolds.fasta.gz
Hydmag	*Hydra magnipapillata*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABRM01.fasta.gz
Hypcat	*Hyphochytrium catenoides*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CAFC02.fasta.gz
Hypsub	*Hypholoma sublateritium*	http://genome.jgi-psf.org/Hypsu1/download/Hypsu1_AssemblyScaffolds.fasta.gz
Hypbur	*Hyphopichia burtonii*	http://genome.jgi-psf.org/Hypbu1/download/Hypbu1_AssemblyScaffolds.fasta.gz
Hyp_sp	*Hypoxylon sp*	http://genome.jgi-psf.org/HypCO275_1/download/HypCO275_1_AssemblyScaffolds.fasta.gz
Ichmul	*Ichthyophthirius multifiliis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEDN01.fasta.gz
Ixosca	*Ixodes scapularis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABJB01.fasta.gz
Jaaarg	*Jaapia argillacea*	http://genome.jgi-psf.org/Jaaar1/download/Jaaar1_AssemblyScaffolds.fasta.gz
Jacjac	*Jaculus jaculus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AKZC01.fasta.gz
Jatcur	*Jatropha curcas*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=BABX01.fasta.gz
Klumar	*Kluyveromyces marxianus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AKFM01.fasta.gz
Kompas	*Komagataella pastoris*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CABH01.fasta.gz
Labfue	*Labeotropheus fuelleborni*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABPK01.fasta.gz
Lacbic	*Laccaria bicolor*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABFE01.fasta.gz
Lacklu	*Lachancea kluyveri*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AACE03.fasta.gz
Lacsat	*Lactuca sativa*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFSA01.fasta.gz
Latcha	*Latimeria chalumnae*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFYH01.fasta.gz
Leeper	*Leersia perrieri*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ALNV01.fasta.gz
Leibra	*Leishmania braziliensis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CADA01.fasta.gz
Lepsal	*Lepeophtheirus salmonis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADND02.fasta.gz
Lepocu	*Lepisosteus oculatus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AHAT01.fasta.gz
Leueri	*Leucoraja erinacea*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AESE01.fasta.gz
Linhum	*Linepithema humile*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADOQ01.fasta.gz
Linusi	*Linum usitatissimum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFSQ01.fasta.gz

Lipsta	*Lipomyces starkeyi*	http://genome.jgi-psf.org/Lipst1_1/download/Lipst1_1_AssemblyScaffolds.fasta.gz
Loaloa	*Loa loa*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADBU02.fasta.gz
Lodelo	*Lodderomyces elongisporus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAPO01.fasta.gz
Lotgig	*Lottia gigantea*	ftp://ftp.jgi-psf.org/pub/JGI_data/Lottia_gigantea/v1.0/Lotgi1_assembly_scaffolds.fasta.gz
Lotjap	*Lotus japonicus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=BABK01.fasta.gz
Loxafr	*Loxodonta africana*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAGU03.fasta.gz
Lutlon	*Lutzomyia longipalpis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AJWK01.fasta.gz
Lytvar	*Lytechinus variegatus*	ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes/Eukaryotes/invertebrates/Lytechinus_variegatus/Lvar_0.4/Pri
Macfas	*Macaca fascicularis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CAEC01.fasta.gz
Macpha	*Macrophomina phaseolina*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AHHD01.fasta.gz
Maceug	*Macropus eugenii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABQO02.fasta.gz
Magory	*Magnaporthe oryzae*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AACU03.fasta.gz
Malglo	*Malassezia globosa*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAYY01.fasta.gz
Maldom	*Malus domestica*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACYM01.fasta.gz
Mansex	*Manduca sexta*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AIXA01.fasta.gz
Manesc	*Manihot esculenta*	ftp://ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/Mesculenta/assembly/Mesculenta_147.fa.gz
Marbru	*Marssonina brunnea*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFXC01.fasta.gz
Maydes	*Mayetiola destructor*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEGA01.fasta.gz
Mayzeb	*Maylandia zebra*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AHAT01.fasta.gz
Mchcon	*Mchenga conophoros*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABPJ01.fasta.gz
Medtru	*Medicago truncatula*	ftp://ftp.jcvi.org/pub/data/m_truncatula/Mt3.5/Assembly/Mt3.5.2/MedtrA17_3.5.assemblies.fasta
Megrot	*Megachile rotundata*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAAB01.fasta.gz
Mellar	*Melampsora larici*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AECX01.fasta.gz
Mel_sp	*Melanconium sp*	http://genome.jgi-psf.org/Melsp1/download/Melsp1_AssemblyScaffolds.fasta.gz
Melaur	*Melanochromis auratus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABPL01.fasta.gz
Melgal	*Meleagris gallopavo*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADDD01.fasta.gz
Melinc	*Meloidogyne incognita*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CABB01.fasta.gz
Melund	*Melopsittacus undulatus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AGAI01.fasta.gz
Menmol	*Mengenilla moldrzyki*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AGDA01.fasta.gz
Metacr	*Metarhizium acridum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADNI01.fasta.gz
Metocc	*Metaseiulus occidentalis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFFJ01.fasta.gz
Metfru	*Metschnikowia fructicola*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ANFW01.fasta.gz
Meygui	*Meyerozyma guilliermondii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAFM01.fasta.gz
Micvio	*Microbotryum violaceum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEIJ01.fasta.gz
Micpus	*Micromonas pusilla*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACCP01.fasta.gz
Micoch	*Microtus ochrogaster*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AHZW01.fasta.gz
Mimgut	*Mimulus guttatus*	ftp://ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/Mguttatus_v1.1/assembly/Mguttatus_140.fa.gz
Mixosm	*Mixia osmundae*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=BABT02.fasta.gz
Mnelei	*Mnemiopsis leidyi*	ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes/Eukaryotes/invertebrates/Mnemiopsis_leidyi/MneLei_Aug201
Monper	*Moniliophthora perniciosa*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABRE01.fasta.gz
Mondom	*Monodelphis domestica*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAFR03.fasta.gz
Monbre	*Monosiga brevicollis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABFJ01.fasta.gz
Morelo	*Mortierella elongata*	http://genome.jgi-psf.org/Morel1/download/Morel1_AssemblyScaffolds.fasta.gz

Muccir	*Mucor circinelloides*	http://genome.jgi-psf.org/Mucci2/download/Mucor_circinelloides_v2_scaffolds.fasta.gz
Musmus	*Mus musculus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEKQ02.fasta.gz
Musacu	*Musa acuminata*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CAIC01.fasta.gz
Musput	*Mustela putorius*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AGTQ01.fasta.gz
Mycthe	*Myceliophthora thermophila*	http://genome.jgi-psf.org/Spoth2/download/Spoth2_AssemblyScaffolds.fasta.gz
Mycpop	*Mycosphaerella populicola*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AIDU01.fasta.gz
Myodav	*Myotis davidii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ALWT01.fasta.gz
Nadful	*Nadsonia fulvescens*	http://genome.jgi-psf.org/Nadfu1/download/Nadfu1_AssemblyScaffolds.fasta.gz
Naegru	*Naegleria gruberi*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACER01.fasta.gz
Nangad	*Nannochloropsis gaditana*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AGNI01.fasta.gz
Nasvit	*Nasonia vitripennis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAZX01.fasta.gz
Naucas	*Naumovozyma castellii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AACF01.fasta.gz
Nechae	*Nectria haematococca*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACJF01.fasta.gz
Nempar	*Nematocida parisii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEFF02.fasta.gz
Nemvec	*Nematostella vectensis*	http://genome.jgi.doe.gov/Nemve1/download/Nemve1.fasta.gz
Neobri	*Neolamprologus brichardi*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFNY01.fasta.gz
Neofis	*Neosartorya fischeri*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAKE03.fasta.gz
Neogan	*Neotyphodium gansuense*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AMDK01.fasta.gz
Neucra	*Neurospora crassa*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AABX02.fasta.gz
Nomleu	*Nomascus leucogenys*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADFV01.fasta.gz
Noscer	*Nosema ceranae*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACOL01.fasta.gz
Ochpri	*Ochotona princeps*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ALIT01.fasta.gz
Octdeg	*Octodon degus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AJSA01.fasta.gz
Ogapar	*Ogataea parapolymorpha*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEOI01.fasta.gz
Oidmai	*Oidiodendron maius*	http://genome.jgi-psf.org/Oidma1/download/Oidma1_AssemblyScaffolds.fasta.gz
Oikdio	*Oikopleura dioica*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CABV01.fasta.gz
Ompole	*Omphalotus olearius*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AHIW01.fasta.gz
Oncvol	*Onchocerca volvulus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADBW01.fasta.gz
Ophcla	*Ophiognomonia clavigignenti*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEGN01.fasta.gz
Orenil	*Oreochromis niloticus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AERX01.fasta.gz
Ornana	*Ornithorhynchus anatinus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAPN01.fasta.gz
Oryafe	*Orycteropus afer*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ALYB01.fasta.gz
Orycun	*Oryctolagus cuniculus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAGW02.fasta.gz
Orysat	*Oryza sativa*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AACV01.fasta.gz
Orylat	*Oryzias latipes*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=BAAF04.fasta.gz
Ostluc	*Ostreococcus lucimarinus*	ftp://ftp.jgi-psf.org/pub/JGI_data/Ostreococcus_lucimarinus/Olucimarinus.fasta.gz
Otogar	*Otolemur garnettii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAQR03.fasta.gz
Oviari	*Ovis aries*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AMGL01.fasta.gz
Oxytri	*Oxytricha trifallax*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AMCR01.fasta.gz
Pactan	*Pachysolen tannophilus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CAHV01.fasta.gz
Pantro	*Pan troglodytes*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AACZ03.fsa.1.gz
Papanu	*Papio anubis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AHZZ01.fsa.1.gz
Parbra	*Paracoccidioides brasiliensis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABKI01.fasta.gz

Partet	*Paramecium tetraurelia*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CAAL01.fasta.gz
Patmin	*Patiria miniata*	ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes/Eukaryotes/invertebrates/Patiria_miniata/Pmin_1.0/Primary_
Paxrub	*Paxillus rubicundulus*	http://genome.jgi-psf.org/Paxru1/download/Paxru1_AssemblyScaffolds.fasta.gz
Pedhum	*Pediculus humanus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAZO01.fasta.gz
Pelsin	*Pelodiscus sinensis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AGCU01.fsa.1.gz
Pendig	*Penicillium digitatum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AKCT01.fasta.gz
Peripo	*Periglandula ipomoeae*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFRD01.fasta.gz
Permar	*Perkinsus marinus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAXJ01.fasta.gz
Petmar	*Petromyzon marinus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEFG01.fasta.gz
Phatri	*Phaeodactylum tricornutum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABQD01.fasta.gz
Phanod	*Phaeosphaeria nodorum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAGI01.fasta.gz
Phacar	*Phanerochaete carnosa*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEHB01.fasta.gz
Phlbre	*Phlebia brevispora*	http://genome.jgi-psf.org/Phlbr1/download/Phlbr1_AssemblyScaffolds.fasta.gz
Phlgig	*Phlebiopsis gigantea*	http://genome.jgi-psf.org/Phlgi1/download/Phlgi1_AssemblyScaffolds.fasta.gz
Phlpap	*Phlebotomus papatasi*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AJVK01.fasta.gz
Phodac	*Phoenix dactylifera*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACYX02.fasta.gz
Phybla	*Phycomyces blakesleeanus*	http://genome.jgi-psf.org/Phybl2/download/Phycomyces_blakesleeanus_v2_scaffolds.fasta.gz
Phypat	*Physcomitrella patens*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABEU01.fasta.gz
Phyinf	*Phytophthora infestans*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AATU01.fasta.gz
Pickud	*Pichia kudriavzevii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ALNQ01.fasta.gz
Pilcro	*Piloderma croceum*	http://genome.jgi-psf.org/Pilcr1/download/Pilcr1_AssemblyScaffolds.fasta.gz
Pinfuc	*Pinctada fucata*	http://marinegenomics.oist.jp/genomes/download/pfu_genome1.0.fasta.gz
Pirind	*Piriformospora indica*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CAFZ01.fasta.gz
Pir_sp	*Piromyces sp*	http://genome.jgi-psf.org/PirE2_1/download/PirE2_1_AssemblyScaffolds.fasta.gz
Pismic	*Pisolithus microcarpus*	http://genome.jgi-psf.org/Pismi1/download/Pismi1_AssemblyScaffolds.fasta.gz
Plaviv	*Plasmodium vivax*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAKM01.fsa
Pleost	*Pleurotus ostreatus*	http://genome.jgi-psf.org/PleosPC15_2/download/PleosPC15_2_Assembly_scaffolds.fasta.gz
Plicri	*Plicaturopsis crispa*	http://genome.jgi-psf.org/Plicr1/download/Plicr1_AssemblyScaffolds.fasta.gz
Pogbar	*Pogonomyrmex barbatus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADIH01.fasta.gz
Polpal	*Polysphondylium pallidum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADBJ01.fasta.gz
Ponabe	*Pongo abelii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABGA01.fsa.1.gz
Poptri	*Populus trichocarpa*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AARH01.fasta.gz
Pospla	*Postia placenta*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABWF01.fasta.gz
Pripac	*Pristionchus pacificus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABKE01.fasta.gz
Procap	*Procavia capensis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABRQ01.fsa.1.gz
Pruper	*Prunus persica*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEJG01.fasta.gz
Psemul	*Pseudo-nitzschia multiseries*	http://genome.jgi-psf.org/Psemu1/download/Psemu1_AssemblyScaffolds.fasta.gz
Psecub	*Pseudoperonospora cubensis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AHJF01.fasta.gz
Ptevam	*Pteropus vampyrus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABRP01.fsa.1.gz
Pucgra	*Puccinia graminis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAWC01.fasta.gz
Punstr	*Punctularia strigosozonata*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEGM01.fasta.gz
Punnye	*Pundamilia nyererei*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFNX01.fasta.gz
Pyrter	*Pyrenophora teres*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEEY01.fasta.gz

Pytult	*Pythium ultimum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADOS01.fasta.gz
Pytmol	*Python molurus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEQU01.fsa.1.gz
Ratnor	*Rattus norvegicus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AABR06.fsa.1.gz
Rhaeso	*Rhamphochromis esox*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABPN01.fasta.gz
Rhopro	*Rhodnius prolixus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACPB02.fasta.gz
Rhotor	*Rhodosporidium toruloides*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AJMJ01.fasta.gz
Rhoglu	*Rhodotorula glutinis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEVR01.fasta.gz
Riccom	*Ricinus communis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AASG02.fasta.gz
Saccer	*Saccharomyces cerevisiae*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACVY01.fasta.gz
Sackow	*Saccoglossus kowalevskii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACQM01.fasta.gz
Saibol	*Saimiri boliviensis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AGCE01.fsa.1.gz
Saicom	*Saitoella complicata*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=BACD01.fasta.gz
Salsal	*Salmo salar*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AGKD01.fasta.gz
Sal_sp	*Salpingoeca sp*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACSY01.fasta.gz
Sappar	*Saprolegnia parasitica*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADCG01.fasta.gz
Sarhar	*Sarcophilus harrisii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEFK01.fsa.1.gz
Schman	*Schistosoma mansoni*	ftp://ftp.sanger.ac.uk/pub/pathogens/Schistosoma/mansoni/genome/Assembly-v5/sma_v5.0.chr.fa.gz
Schcom	*Schizophyllum commune*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADMJ01.fasta.gz
Schjap	*Schizosaccharomyces japonicus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AATM01.fasta.gz
Schmed	*Schmidtea mediterranea*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAWT01.fasta.gz
Sclcit	*Scleroderma citrinum*	http://genome.jgi-psf.org/Sclci1/download/Sclci1_AssemblyScaffolds.fasta.gz
Sclscl	*Sclerotinia sclerotiorum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAGT01.fasta.gz
Sebver	*Sebacina vermifera*	http://genome.jgi-psf.org/Sebve1/download/Sebve1_AssemblyScaffolds.fasta.gz
Selmoe	*Selaginella moellendorffii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADFJ01.fasta.gz
Serlac	*Serpula lacrymans*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEQB01.fasta.gz
Setita	*Setaria italica*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AGNK01.fasta.gz
Settur	*Setosphaeria turcica*	http://genome.jgi-psf.org/Settu1/download/Settu1_AssemblyScaffolds.fasta.gz
Sollyc	*Solanum lycopersicum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEKE02.fasta.gz
Solinv	*Solenopsis invicta*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEAQ01.fasta.gz
Sormac	*Sordaria macrospora*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CABT02.fasta.gz
Sorara	*Sorex araneus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AALT02.fsa.1.gz
Sorbic	*Sorghum bicolor*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABXC01.fasta.gz
Spapas	*Spathaspora passalidarum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEIK01.fasta.gz
Spetri	*Spermophilus tridecemlineatus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AGTP01.fsa.1.gz
Sphste	*Sphaerobolus stellatus*	http://genome.jgi-psf.org/Sphst1/download/Sphst1_AssemblyScaffolds.fasta.gz
Spipun	*Spizellomyces punctatus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACOE01.fasta.gz
Sporos	*Sporobolomyces roseus*	ftp://ftp.jgi-psf.org/pub/JGI_data/Sporobolomyces_roseus/assembly/v1.0/Sporobolomyces_roseus.allmas
Stehir	*Stereum hirsutum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEGX01.fasta.gz
Strmar	*Strigamia maritima*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFFK01.fasta.gz
Strpur	*Strongylocentrotus purpuratus*	ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes/Eukaryotes/invertebrates/Strongylocentrotus_purpuratus/Spu
Strrat	*Strongyloides ratti*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CACX01.fasta.gz
Stylem	*Stylonychia lemnae*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADNZ01.fasta.gz
Suilut	*Suillus luteus*	http://genome.jgi-psf.org/Suilu1/download/Suilu1_AssemblyScaffolds.fasta.gz

Susscr	*Sus scrofa*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEMK01.fsa.1.gz
Taegut	*Taeniopygia guttata*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABQF01.fsa.1.gz
Takrub	*Takifugu rubripes*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CAAB02.fasta.gz
Talmar	*Talaromyces marneffei*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABAR01.fasta.gz
Tapdef	*Taphrina deformans*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CAHR02.fasta.gz
Tarsyr	*Tarsius syrichta*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABRT01.fsa.1.gz
Tetthe	*Tetrahymena thermophila*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAGF03.fasta.gz
Teturt	*Tetranychus urticae*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CAEY01.fasta.gz
Tetnig	*Tetraodon nigroviridis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CAAE01.fasta.gz
Thapse	*Thalassiosira pseudonana*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAFD02.fasta.gz
Thepar	*Theileria parva*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAGK01.fasta.gz
Thecac	*Theobroma cacao*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CACC01.fasta.gz
Thelan	*Thermomyces lanuginosus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ANHP01.fasta.gz
Thiter	*Thielavia terrestris*	http://genome.jgi-psf.org/Thite2/download/Thite2_AssemblyScaffolds.fasta.gz
Toxgon	*Toxoplasma gondii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABPA01.fasta.gz
Traver	*Trametes versicolor*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEJI01.fasta.gz
Tremes	*Tremella mesenterica*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFVY01.fasta.gz
Tricas	*Tribolium castaneum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAJJ01.fasta.gz
Triman	*Trichechus manatus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AHIN01.fsa.1.gz
Trispi	*Trichinella spiralis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABIR02.fasta.gz
Triree	*Trichoderma reesei*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAIL02.fasta.gz
Trivag	*Trichomonas vaginalis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAHC01.fasta.gz
Trirub	*Trichophyton rubrum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACPH01.fasta.gz
Triadh	*Trichoplax adhaerens*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABGP01.fasta.gz
Triasa	*Trichosporon asahii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AMBO01.fasta.gz
Triaes	*Triticum aestivum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEOM01.fasta.gz
Trycru	*Trypanosoma cruzi*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAHK01.fasta.gz
Tubmel	*Tuber melanosporum*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CABJ01.fasta.gz
Tulcal	*Tulasnella calospora*	http://genome.jgi-psf.org/Tulca1/download/Tulca1_AssemblyScaffolds.fasta.gz
Tupbel	*Tupaia belangeri*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAPY01.fsa.1.gz
Turtru	*Tursiops truncatus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABRN02.fsa.1.gz
Uncree	*Uncinocarpus reesii*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAIW01.fasta.gz
Usthor	*Ustilago hordei*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CAGI01.fasta.gz
Vanpol	*Vanderwaltozyma polyspora*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAZN01.fasta.gz
Vardes	*Varroa destructor*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADDG01.fasta.gz
Vavcul	*Vavraia culicis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEUG01.fasta.gz
Verdah	*Verticillium dahliae*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABJE01.fasta.gz
Vicpac	*Vicugna pacos*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ABRR01.fsa.1.gz
Vitvin	*Vitis vinifera*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=CAAP03.fasta.gz
Vitcor	*Vittaforma corneae*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEYK01.fasta.gz
Volcar	*Volvox carteri*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ACJH01.fasta.gz
Walseb	*Wallemia sebi*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFQX01.fasta.gz
Wicano	*Wickerhamomyces anomalus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AEGI02.fasta.gz

Wolcoc	*Wolfiporia cocos*	http://genome.jgi-psf.org/Wolco1/download/Wolco1_AssemblyScaffolds.fasta.gz
Wucban	*Wuchereria bancrofti*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=ADBV01.fasta.gz
Xanpar	*Xanthoria parietina*	http://genome.jgi-psf.org/Xanpa1/download/Xanpa1_AssemblyScaffolds.fasta.gz
Xentro	*Xenopus tropicalis*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AAMC02.fsa.1.gz
Xipmac	*Xiphophorus maculatus*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AGAJ01.fasta.gz
Zascel	*Zasmidium cellare*	http://genome.jgi-psf.org/Zasce1/download/Zasce1_AssemblyScaffolds.fasta.gz
Zeamay	*Zea mays*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AECO01.fasta.gz
Zymard	*Zymoseptoria ardabiliae*	http://www.ncbi.nlm.nih.gov/Traces/wgs/?download=AFIV01.fasta.gz

CHAPTER 3

Evolutionary and biogeographical patterns of barnacles from deep-sea hydrothermal vents

Santiago Herrera [1,2], Hiromi Watanabe[3], Timothy M. Shank [2]

[1] Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

[2] Biology Department, Woods Hole Oceanographic Institution, 266 Woods Hole Road, Woods Hole, MA 02543, USA

[3] Institute of Biogeosciences, Japan Agency for Marine-Earth Science and Technology, Yokosuka, Kanagawa, Japan

ABSTRACT

The characterization of evolutionary and biogeographical patterns is of fundamental importance to identify factors driving biodiversity. Due to their widespread but discontinuous distribution, deep-sea hydrothermal vent barnacles represent an excellent model for testing biogeographic hypotheses regarding the origin, dispersal, and diversity of modern vent fauna. Here we characterize the global genetic diversity of vent barnacles to infer their time of radiation, place of origin, mode of dispersal, and diversification. Our approach was to target a suite of multiple loci in samples representing seven out of the eight described genera. We also performed restriction-site associated DNA sequencing on individuals from each species. Phylogenetic inferences and topology hypothesis tests indicate that vent barnacles have colonized deep-sea hydrothermal vents at least twice in history. Consistent with preliminary estimates, we find a likely radiation of barnacles in vent ecosystems during the Cenozoic. Our analyses suggest that the western Pacific was the place of origin of the major vent barnacle lineage, followed by circumglobal colonization eastward through the southern hemisphere during the Neogene. The inferred time of radiation rejects classic hypotheses of antiquity of vent taxa. The timing and the mode of origin, radiation and dispersal are consistent with recent inferences made for other deep-sea taxa, including non-vent species, and are correlated with the occurrence of major geological events and mass extinctions. Thus, we suggest that the geological processes and dispersal mechanisms discussed here can explain current distribution patterns of many other marine taxa and have played an important role shaping deep-sea

faunal diversity. These results also constitute critical baseline data with which to assess potential effects of anthropogenic disturbances on deep-sea ecosystems.

INTRODUCTION

The characterization of evolutionary and biogeographical patterns is of fundamental importance for identifying the factors that shape the ranges of deep-sea taxa, and that ultimately drive biodiversity patterns in the ocean (McClain & Mincks 2010). This is particularly relevant in the light of the increasing interest in commercial resource extraction in the deep-sea (Thurber *et al.* 2014). Mining of seafloor massive sulphide deposits at deep-sea hydrothermal vent fields has become one of the main industrial targets for exploitation (Boschen *et al.* 2013). Understanding the biodiversity contained in these areas and its connection with the fauna found elsewhere is critical for assessing the potential impacts of exploiting these mineral resources (Van Dover 2010; Van Dover *et al.* 2012). Although organisms living at deep-sea hydrothermal vents have adapted to cope with natural disturbances inherent to these ephemeral habitats, the intensity and frequency at which these occur can vary greatly depending on the particular geophysical nature of each system (Baker & German 2004). Thus, disturbance from mining could have additive or synergistic effects to natural disturbances at unprecedented scales, which could potentially lead to significant losses of biodiversity (Van Dover 2010). Due to their widespread distribution (Fig. 1), vent barnacles represent an excellent model for testing hypotheses regarding the historical biogeographic patterns of origin, dispersal, and current diversity of modern deep-sea chemosynthetic fauna; therefore, barnacles hold the promise of providing critical baseline data with which to assess potential effects of anthropogenic disturbances on deep-sea ecosystems.

Barnacles (Cirripedia Burmeister, 1834) are some of the most conspicuous organisms in deep-sea hydrothermal vent ecosystems worldwide. These sessile crustaceans can be found in active vent fields in most of the major spreading ridge systems and island arcs worldwide (Fig. 1), including the Central Indian Ridge (Van Dover *et al.* 2001; Nakamura *et al.* 2012), Southwest Indian Ridge (Tao *et al.* 2011), East Scotia Ridge (Rogers *et al.* 2012), northern and southern East Pacific Rise (Newman 1979; Jones 1993), Pacific-Antarctic Ridge (Southward 2005), Izu-Ogasawara Arc (Ohno *et al.* 1996), Okinawa Trough (Ohta 1990), Mariana Trough (Hessler & Lonsdale 1991), Sangihe Talaud (Herrera *et al.* 2010; Shank *et al.* 2010), Manus Basin (Tufar 1990), Edison Seamount (Tunnicliffe & Southward 2004), North Fiji Basin (Desbruyeres *et al.* 1994), Lau Basin (Southward & Newman 1998), Kermadec Arc (Buckeridge 2000), and are likely to be present in other unexplored areas. Hydrothermal vent barnacles inhabit areas of low-temperature diffuse fluid flow. Populations can reach high densities with high

biomass at over 1500 individuals per square meters (Tunnicliffe & Southward 2004; Marsh *et al.* 2012), playing key roles in vent communities as micro-habitat engineers and funnelling the flow of energy through ecosystems from primary producers to higher trophic levels (Southward & Newman 1998; Van Dover 2002; Tunnicliffe & Southward 2004; Cubelio *et al.* 2007; Rogers *et al.* 2012; Reid *et al.* 2013).

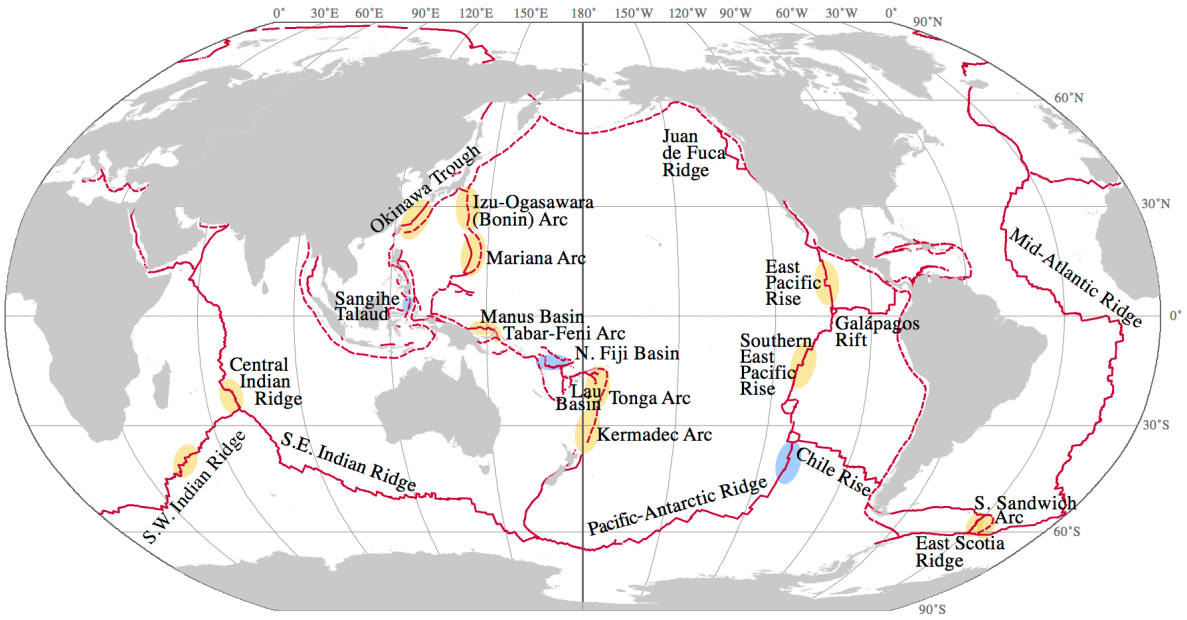


Figure 1. Global distribution map of hydrothermal vent barnacles. Ovals indicate regions where hydrothermal vent barnacles have been described (yellow: regions sampled in this study; blue: regions not sampled in this study). Red lines indicate active tectonic margins (solid lines: spreading centers; dotted lines: subduction zones).

Hydrothermal vent barnacles are presently grouped into four families belonging to the orders Pedunculata Lamarck, 1818 (suborder Scalpellomorpha, family Eolepadidae; commonly known as stalked or gooseneck barnacles) and Sessilia Lamarck, 1818 (suborder Verucomorpha, family Neoverrucidae; suborder Brachylepadomorpha, family Neobrachylepadidae; and suborder Balanomorpha, family Chionelasmatidae; commonly known as acorn barnacles) (Newman *et al.* 2006). There are approximately 13 described vent barnacle species, with several new species awaiting description (Newman *et al.* 2006). A molecular phylogenetic study of the Cirripedia, employing nuclear ribosomal genes and the histone *H3* gene, indicates that these morphologically-based taxonomic groupings (orders) are polyphyletic and thus incongruent with evolutionary history (Pérez-Losada *et al.* 2008). These results, together with those from (Linse *et al.* 2013), also suggest that vent barnacles form a monophyletic clade that likely originated in the Cretaceous; however, the possibility of a single origin remains an open question due to the paucity of

taxonomic sampling in that study. Furthermore, the relationships among morphospecies of vent barnacles also remain unresolved due to the low variability of markers examined to date.

Many putative species of vent barnacles appear to be restricted to particular ridge systems and neighboring arc and back-arc basins, and significant population structure has also been found at these scales (Watanabe *et al.* 2005). Together these observations suggest a role of habitat discontinuity as an important mechanism of speciation. By far, the region of highest diversity of putative chemosynthetic barnacle species (measured as species richness) is the western Pacific, which is considered the center of their distribution and possible place of origin (Newman *et al.* 2006). The western Pacific is also considered a biodiversity hotspot and potential place of origin of many modern groups of terrestrial and marine organisms, including deep-sea taxa (Cairns 2007; Carpenter *et al.* 2011; Herrera *et al.* 2012). In a similar way, a recent biogeographic analyses using network theory hypothesizes a possible ancestral position of the western Pacific for modern fauna associated with hydrothermal vents, having exclusive edge connections (indicating faunal similarity possible exchange paths) with the Northeast Pacific, the East Pacific Rise and the Indian Ocean (Moalic *et al.* 2011).

In this study, we aim to characterize the global genetic diversity, evolutionary and biogeographic history of barnacles from deep-sea hydrothermal vents. Our approach was to build on previous phylogenetic studies by significantly expanding the taxonomic sampling and number of genetic markers. We targeted one mitochondrial gene region, the cytochrome c oxidase subunit I (*coxI*), and two nuclear gene regions, the large ribosomal sub-unit 28S, and the histone *H3* gene, obtaining complete sequences for 94 individuals, representing seven out of the eight described genera, from 18 vent fields worldwide. We also performed restriction-site associated DNA sequencing (RAD-seq) on individuals from each identified species. Here we: (1) test the hypothesis of monophyly (i.e., a single evolutionary origin) of barnacles from deep-sea hydrothermal vents; (2) infer the place and time of origin and radiation of vent barnacles in geologic time; (3) infer historical patterns of dispersal and colonization of vent barnacle taxa worldwide; and (4) identify species boundaries and compare them to current morphospecies hypotheses.

METHODS

Morphological identifications were performed on 94 barnacle specimens (Table S1) from deep-sea hydrothermal vents using stereo-microscopy and species descriptions as references. Individuals were collected from the Central Indian Ridge, East Pacific Rise, southern East Pacific Rise, Southwest Indian

Ridge, East Scotia Ridge, Mariana Trough, the Kermadec Arc, Lau Basin, Tonga Arc, Manus Basin, Izu-Ogasawara (Bonin) Arc, and the Okinawa Trough.

Partial DNA sequences of one mitochondrial (cytochrome c oxidase subunit I) and two nuclear markers (histone *H3* gene and the ribosomal large sub-unit 28S) were generated for each individual. Additional sequences from the Superorder Thoracica Darwin, 1854 were retrieved from GenBank (http://www.ncbi.nlm.nih.gov/genbank/) and included in the analyses (Table S2).

Restriction-site Associated DNA sequencing (RAD-seq) (Baird *et al.* 2008) was performed on selected individuals from each morphospecies (Table S1) to obtain a genome-wide set of markers that could be used to infer a robust backbone of the vent barnacle phylogenetic tree, and to compare to topologies obtained from species-tree analyses of traditional Sanger-based markers.

Molecular laboratory methods

Total genomic DNA was extracted from tissue samples by: (1) digesting the tissue in 2 % CTAB buffer (Teknova) with proteinase K and RNAse A/T1 (Fermentas) for 1 hour, (2) separating nucleic acids with chloroform: isoamyl alcohol (24:1) (Fermentas) and phenol: chloroform: isoamyl alcohol (25:24:1, Tris buffered at pH 8.0) (Fermentas), (3) precipitating nucleic acids with 100% ethanol (1:1), and (4) washing the precipitate twice with 70% ethanol. Polymerase chain reactions of traditional Sanger-based markers were prepared to a final volume of 25 µl (1 µl of template) resulting in the following final concentrations of reagents and enzymes: 1 X GoTaq Flexi Buffer (Promega), 2.5 X BSA, 1.0 mM dNTPs (0.25 mM each), 2.0 mM $MgCl_2$, 1 U Taq polymerase (GoTaq, Promega), and 0.3 µM of each primer. Primer pairs used for amplifications were: 28SF_330 5'- CGTGAAGCTGCCAVTATGG-3' (designed in this study) & 28S_B (Whiting 2002) for 28S, H3F & H3R (Colgan *et al.* 1998) for *H3*, and LCO1490 & HC02198 (Folmer *et al.* 1994) for *coxI*. Negative controls were included in every experiment to test for contamination. The reactions were carried out with an initial denaturation step of 5 min at 94 °C, 32 cycles (35 for *coxI*) of 60 s at 94 °C, 90 s at 48 °C, and 90 s at 72 °C, and a final elongation step of 10 min at 72 °C. PCR products were cleaned using the MinElute PCR Purification Kit (Qiagen) following manufacturer's protocols. Cycle-sequencing reactions were performed using the ABI BigDye Terminator v3.1 kit (Life Technologies Corp.), following manufacturer protocols. Subsequent purification was done through isopropanol precipitation. Automated sequencing was completed using a 3730xl DNA analyzer (Life technologies Corp.) at the Josephine Bay Paul Center of the Marine Biological Laboratory. Complementary chromatograms were assembled and edited using Geneious v6.1.6 (Drummond *et al.* 2011).

Concentration-normalized genomic DNA was submitted to Floragenex Inc. (Eugene, OR) for library preparation and RAD sequencing. Individual libraries were produced from DNA digested with a high-fidelity SbfI restriction enzyme, which is predicted to cut approximately between 5,000 and 15,000 times in the genome of a thoracican barnacle (Table S3). This predicted range was obtained using the observed frequency of the SbfI recognition sequence, and its probability using a trinucleotide composition model, in the genome of the crustacean *Daphnia pulex* (Herrera *et al*. Chapter 2). Ranges of genome size for barnacles were obtained from the Animal Genome Size Database (http://ww.genomesize.com). Barcode tags were 10-base pairs long. Libraries were sequenced by 96-multiplex on a single lane of an Illumina Hi-Seq 2000 sequencer.

Alignments, saturation analysis and model selection

Each set of sequences for Sanger-based markers was aligned independently using MAFFT (Katoh *et al.* 2002), employing the G-INS-i and Q-INS-i algorithms (gap opening penalty= 1.53, offset value= 0.07) for protein coding and ribosomal regions, respectively. To correct possible mistakes, all alignments of protein coding sequences were visually inspected and translated to amino acids in Geneious v6.1.6 (Drummond *et al.* 2011). No unusual stop codons, misplaced reading frames or suspicious substitutions were identified, indicating amplification of nuclear pseudogenes was unlikely (Lopez *et al.* 1994; Bensasson *et al.* 2001). Possible substitution saturation in the DNA sequences was evaluated by implementing the Xia test (Xia *et al.* 2003), as implemented in DAMBE v5.3.48 (Xia 2013), and by plotting genetic distances (K80 model) against the number of transitions and transversions. Saturation in codon partitions was also evaluated for each coding region.

Phylogenetic Inferences

Non-saturated datasets from individual Sanger-based markers were analyzed in RAxML-HPC2 v8.0 (Stamatakis 2006), as implemented in the CIPRES Science Gateway v3.3 (http://www.phylo.org), for a first-pass phylogenetic inference using the maximum likelihood optimality criterion. Branch support was assessed by 500 bootstrap replicates. A Thoracica-wide concatenated dataset was also analysed in this program. Only outgroups with data for at least two of the three markers were included in the concatenated dataset. Phylogenetic estimation through Bayesian inference was performed with these datasets in MrBayes v3.2.2 (Ronquist *et al.* 2012), as implemented in the CIPRES Science Gateway v3.3. Models of nucleotide substitution were selected for each non-saturated gene region using JModeltest v2.0 (Darriba *et al.* 2012), following the Bayesian Information Criterion (Table S4). Four independent analyses of 200 million Markov Chain Monte Carlo (MCMC) generations (4 chains) were run with a sampling frequency

of 20 thousand generations (burn-in = 25%). Combined analyses were performed with explicit character partitions for each concatenated region, along with their independently selected models of evolution. State frequencies were allowed to vary under a flat Dirichlet prior distribution to account for the rate variation among partitions. Nucleotide frequencies, substitution rates, gamma shape and invariant site proportion parameters were unlinked across partitions. Default prior distribution settings were assumed for all other parameters. MCMC runs were analyzed with the programs Tracer v1.5 (Rambaut & Drummond 2007) and AWTY (http://ceb.csit.fsu.edu/awty) (Nylander *et al.* 2008). Convergence among independent runs was supported by observed values of standard deviation of partition frequencies (<0.01), potential scale reduction factors (PSRF) (ca. 1.00), and effective sample sizes (EES) (>200), in addition to high correlations between runs and the flat shapes of the stationary posterior distribution traces of each parameter.

Topological Hypothesis Testing

To test the hypothesis that barnacles from deep-sea hydrothermal vents form a monophyletic group, we performed a Bayes factor comparison (Kass & Raftery 1995) between this topological hypothesis and the alternative hypothesis of non-monophyly of the group using the Thoracica-wide concatenated dataset. The marginal likelihood for each topology model was estimated through the stepping-stone method (Fan *et al.* 2011; Xie *et al.* 2011) in MrBayes using 50 steps. The estimation was performed in two independent runs of 100 million generations, with a diagnostic frequency of 1 million generations, for each topology model. All other parameters were set to default. Convergence among runs was diagnosed by the standard deviation of partition frequencies (<0.01).

Divergence Time Estimations

Time-calibration of the phylogenetic hypothesis was carried out through a Bayesian-MCMC joint estimation of phylogeny and divergence times in BEAST v1.7.5 (Drummond *et al.* 2012), as implemented in the CIPRES Science Gateway v3.3, using the Thoracica-wide concatenated Sanger-based markers dataset. Variation in mutation rates among branches was allowed by assuming an uncorrelated relaxed lognormal molecular clock model. The Yule constant speciation rate model and no extinction (Yule 1925), the Birth-Death constant speciation and extinction rates model (Gernhard 2008), and the Birth-Death constant speciation and extinction rates with incomplete taxonomic sampling model (Stadler 2009) were tested as tree priors. Unlinked character partitions were set for each concatenated region, along with their independently selected models of evolution. Three fossil calibration points (C1, C2, and C7) were selected from the studies by Pérez-Losada *et al.* (2008) and Linse *et al.* (2013) based on well-supported topological congruencies with our phylogenetic hypothesis. Fossil ages were used as lower boundary

constraints assuming prior exponential distributions with mean values of 25 my. Default prior distribution settings were assumed for all other parameters. Three independent MCMC analyses were run for 200 million generations with a sampling frequency of 20 thousand. Convergence diagnostics were examined for the combined runs in Tracer as mentioned above. Most probable trees, after 25% burn-in, were summarized into a maximum clade credibility tree with median node heights using TreeAnnotator v1.7.1 (Drummond *et al.* 2012).

Historical Biogeography

To infer historical patterns of dispersal in deep-sea hydrothermal vent barnacle lineages, we performed a Bayesian reconstruction of discrete character states of geographic location for ancestral nodes (Lemey *et al.* 2009) using BEAST v1.7.5 (Heled & Drummond 2010). In this framework, the geographical sampling locations were mapped to the time-scaled phylogenetic tree. Parameters for tree inference were as described above.

Species Delimitation

To identify species boundaries for vent barnacles in Clade A (see Results section), we employed generalized mixed Yule-coalescent (GMYC) likelihood method (Pons *et al.* 2006; Monaghan *et al.* 2009; Fujisawa & Barraclough 2013), with a single threshold, as implemented in the SPLITS R-package (available from http://r-forge.r-project.org/ projects/splits/). This method estimates species boundaries by identifying increases in branching rates that are characteristic of transition points between interspecific speciation–extinction processes and intraspecific coalescent processes.

Species Tree Estimation

Bayesian analyses of species-trees estimation for vent barnacle species identified in Clade A (see Results section) were carried out using data from the Sanger-based markers in the program *BEAST v1.7.5 (Heled & Drummond 2010). This approach was employed to take into account evolutionary coalescent processes and gene tree heterogeneity, and to evaluate the effects of gene-concatenation on the phylogenetic inference (Brito & Edwards 2008; Edwards 2008). Species were defined after the species delimitation analyses. Unlinked character, clock, and tree partitions were set for each marker, along with their independently selected models of evolution. We assumed a piecewise linear and constant root population size model. Other parameters for tree inference were as described above.

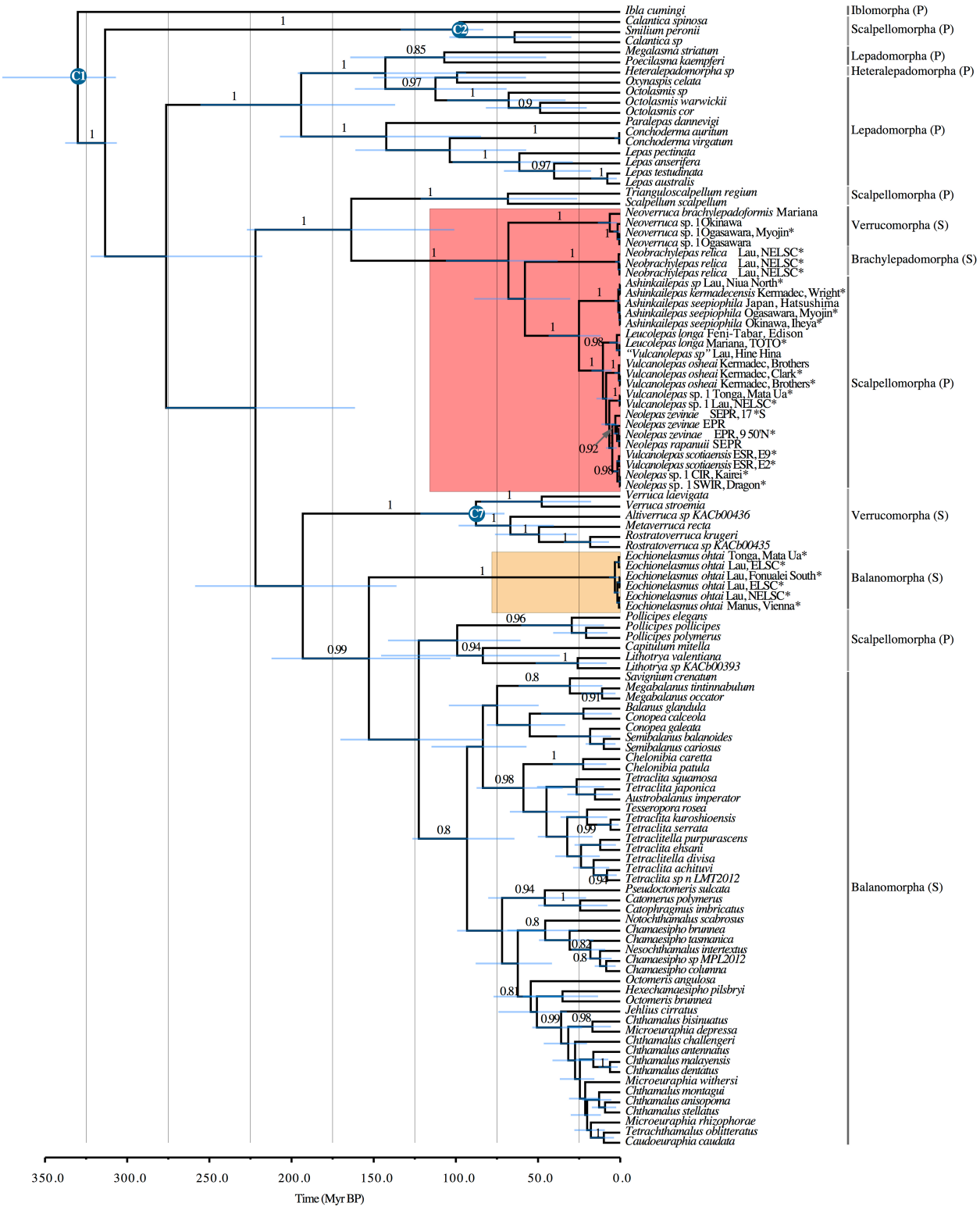


Figure 2. Maximum clade credibility ultrametric time-scaled tree, generated under the Birth-Death model tree prior, for the Thoracica-wide concatenated dataset. Red square indicates hydrothermal vent Clade A. Yellow square

indicates hydrothermal vent Clade B. Node bars represent the 95% highest posterior density intervals. Branch labels show posterior probabilities. Blue circles in nodes indicate fossil calibration points as in (Pérez-Losada *et al.* 2008; Linse *et al.* 2013). Suborders belonging to the order Pedunculata (stalked or gooseneck barnacles) are indicated with (P). Suborders belonging to the order Sessilia (acorn barnacles) are indicated with (S). *Indicates data generated in this study.

RAD-seq data quality control and loci clustering

Sequence reads were de-multiplexed and quality filtered with the process_radtags program from the package Stacks v1.19 (Catchen *et al.* 2011; Catchen *et al.* 2013). Barcodes and Illumina adapters were excluded from each read and length was truncated to 90bp (-t 90) Reads with ambiguous bases were discarded (-c). Reads with an average quality score below 10 (-s 10) within a sliding window of 15% of the read length (-w 0.15) were discarded (-r). The rescue barcodes and RAD-tags algorithm was enabled (-r). Additional filtering, and the clustering within and between individuals to identify loci was performed using the program *py*RAD v2.01 (Eaton 2014). Reads with more than 33 bases with a quality score below 20 were also discarded. The minimum depth of coverage required to build a cluster was 5 (d 5). As in Hipp *et al.* (2014), three different clustering thresholds were explored (c 0.80, 0.85 and 0.90). Similarly, four different values for the minimum taxon coverage in a given locus were explored (m 4, 6, 8, and 10). The maximum number of shared polymorphic sites in a locus was set to 3 (p 3). Loci were concatenated into combined RAD-seq matrices.

RAD Phylogenetics

Phylogenetic inferences of RAD-seq matrices, built with *py*RAD under each combination of clustering threshold and minimum taxon coverage parameters (as outlined above), were carried out in RAxML-HPC2 v8.0. We assumed a generalized time-reversible DNA substitution model with a gamma-distributed rate variation across sites. Branch support was assessed by 500 bootstrap replicates.

RESULTS

Complete Sanger-based marker datasets were obtained for all 94 individuals, except for 2 specimens of *Vulcanolepas osheai.* Sequences are stored in the GenBank database of the U.S. National Center for Biotechnology Information. Approximate sequence lengths for each marker were 700 bp for 28S, 657 bp for *coxI,* and 327 bp for *H3*. Xia tests indicated substantial saturation at the Thoracica-wide level at third codon positions of *coxI* (Table S5). Little saturation was found in all other partitions. Maximum likelihood and Bayesian phylogenetic inferences from each Sanger-based marker produced mostly congruent trees that varied in the degree of resolution yet all showed poorly supported branches (i.e.,

posterior probability < 80, bootstrap support < 80) (supplementary electronic material). Analyses of the Thoracica-wide concatenated dataset generated a better-supported and -resolved phylogeny overall (Fig. 2, supplementary electronic material). The topologies of these trees were congruent with previously published phylogenetic hypotheses for the Thoracica (Pérez-Losada *et al.* 2008; Linse *et al.* 2013).

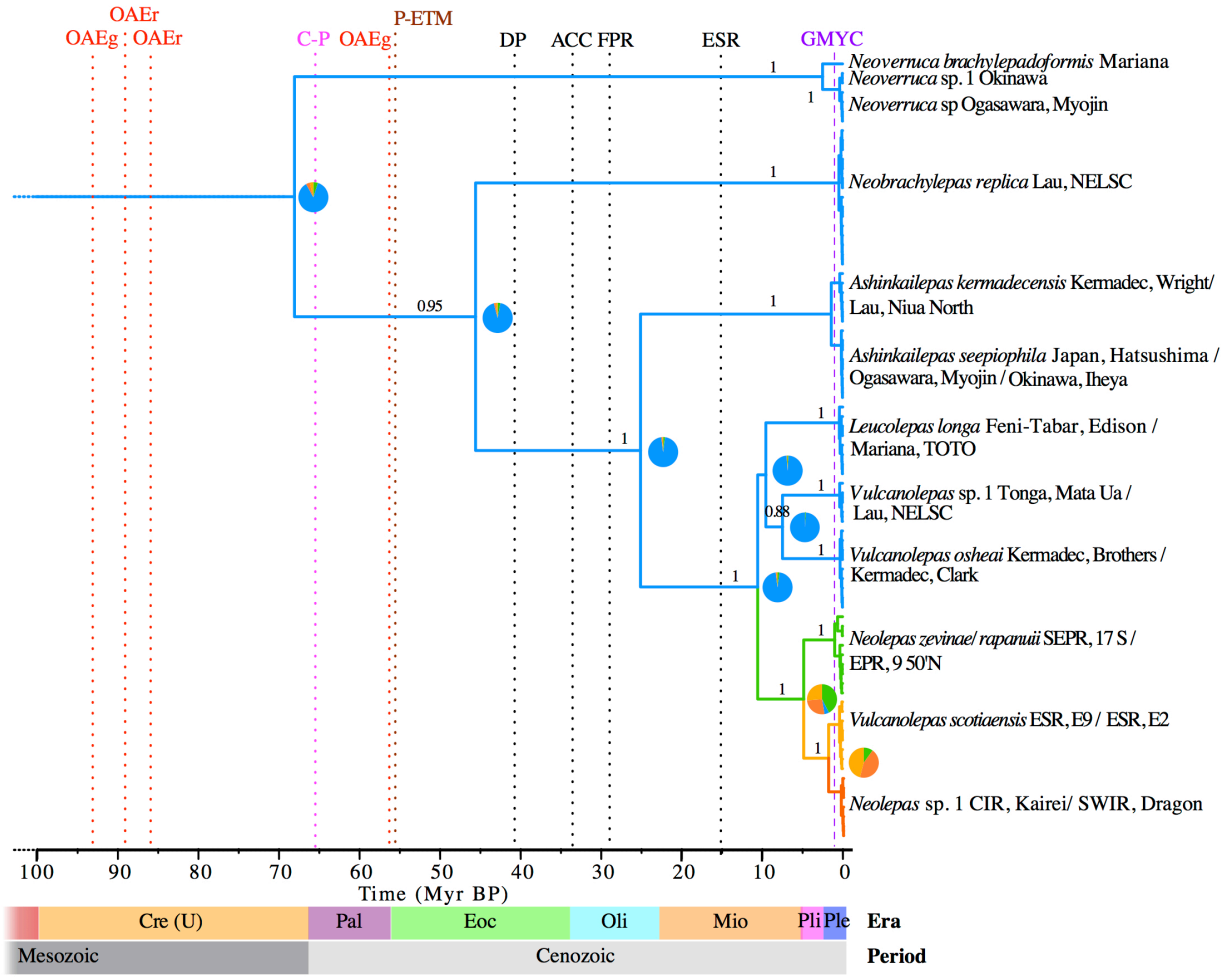


Figure 3. Clade A combined 28S, *H3*, and *coxI* maximum clade credibility ultrametric time-scaled tree generated under the Birth-Death model. Branch colours show the most probable location states: western Pacific in blue, eastern Pacific in green, Southern Ocean south of the Atlantic in yellow and Indian Ocean in orange. Pie charts show the posterior probabilities of location states for each ancestral node (total pie area = 1). Branch labels show posterior probabilities. Purple vertical dashed line indicates the maximum likelihood-inferred time for the speciation-coalescent threshold for species delimitation (GYMC). Vertical dotted lines indicate important events in geologic time: Oceanic Anoxic Events (red, OAEg for global and OAEr for regional), Cretaceous-Paleogene mass extinction

(fuchsia, C-P), Paleocene-Eocene Thermal Maximum (brown, P-ETM), opening of the Drake Passage (black, DP), establishment of the Antarctic Circumpolar Current (black, ACC), disruption of the Farallon Pacific Ridge (black, FPR), formation of the East Scotia Rise (black, ESR). Geologic periods and eras are indicated with horizontal bars: upper Cretaceous (Cre (U)), Paleocene (Pal), Eocene (Eoc), Oligocene (Oli), Miocene (Mio), Pliocene (Pli), and Pleistocene (Ple).

RAD-seq datasets were obtained from 13 individuals representing the vent barnacle species in this study (Table S1). An average of 843,541 reads (SD 589,377) were obtained per individual. Reads are stored at the Sequence Read Archive (SRA) of NCBI. The great variability in sequencing yield was largely a product of varying of DNA integrity as some samples had notably degraded DNA (Table S6), as determined by agarose gel electrophoresis. An average of 712,306 reads per individual (SD 546,846), roughly 78% of all reads, were retained after quality filtering steps. In individuals with high-integrity DNA, the number of RAD-tag loci with depth of coverage greater than 4 X was approximately 18,000, per individual. This number is congruent with the expected number of RAD-tags, between 10,000 and 30,000, predicted for a barnacle, using the enzyme SbfI (Table S3). The average depth of coverage per locus was approximately 54 X (SD 13 X). As expected, the number of loci per individual was higher as the clustering threshold was larger (Table S7). Phylogenetic trees obtained from the RAD-seq datasets were completely resolved, highly supported as indicated by bootstrap resampling, and were largely congruent with the trees produced with Sanger-based data.

Phylogenetic Inferences

Analyses of Sanger-based markers revealed that barnacles from deep-sea hydrothermal vents are divided into two well-supported (posterior probability = 1, bootstrap support > 99) main clades (Fig. 2): Clade A contains the genera *Neobrachylepas* Newman & Yamaguchi, 1995 (Order Sessilia, Suborder Brachylepadomorpha), *Neoverruca* Newman, 1989 (Order Sessilia, Suborder Verrucomorpha), *Ashinkailepas* Yamaguchi *el al.* 2004 (Order Pedunculata, Suborder Scalpellomorpha), *Leucolepas* Southward & Jones, 2003 (Suborder Scalpellomorpha), *Vulcanolepas* Southward & Jones, 2003 (Suborder Scalpellomorpha), and *Neolepas* Newman, 1979 (Suborder Scalpellomorpha); and Clade B was restricted to the genus *Eochionelasmus* Yamaguchi, 1990 (Order Sessilia, Suborder Balanomorpha). Clade A is well supported as the sister taxon to the predominantly deep-sea clade of the Scalpellidae (Pérez-Losada *et al.* 2008; Linse *et al.* 2013). Clade B *Eochionelasmus* is associated with the paraphyletic Balanomorpha group, however the lack of support and resolution within the later group prevents an unambiguous phylogenetic placement.

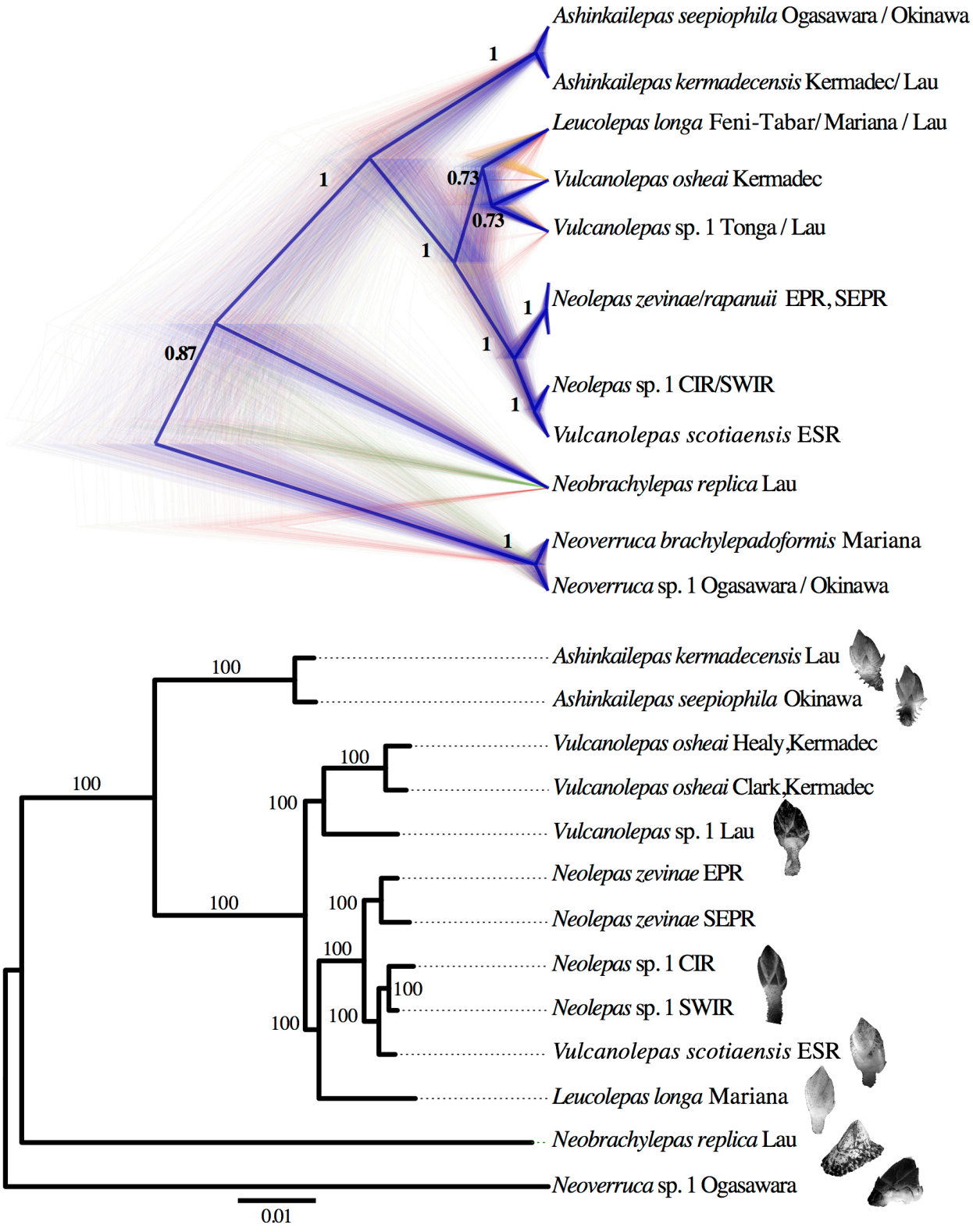


Figure 4. *Top*. Claudogram of the posterior distribution of species trees. High color density is indicative of areas in the species trees with high topology agreement. Different colors represent different topologies. The maximum clade

credibility species tree is shown with thicker branches. Branch labels show posterior probabilities. Trees with the same topology as the maximum clade credibility species tree are coloured in blue. Trees with different topologies are colored yellow or red. *Bottom*. Maximum likelihood phylogenetic tree inferred with RAD-seq data. The matrix used for this tree was obtained with a clustering threshold of 0.85 and minimum taxon coverage of 6. This matrix contains 828,960 nucleotide sites in 9,766 loci. 76,353 of the sites are variable and 26,955 are parsimony informative. This matrix contains 43.54% missing data. Branch labels show bootstrap support values. Scale bar indicates substitutions per site. Barnacle species images are from individuals included in this study. Species names are followed by the collection regions.

Neobrachylepas and *Neoverruca* appear as the extant representatives of the earliest divergent lineages in Clade A; however, their order of divergence is unclear due to lack of strong branch support. The rest of the genera in Clade A belong to the Family Eolepadidae. The genus *Ashinkailepas* belongs to the earliest divergent lineage in the family (Fig. 3), and contains two sub-clades, one grouping individuals from the Izu-Ogasawara (Bonin) Arc and the Okinawa Trough (identified as *Ashinkailepas seepiophila*), and the second grouping individuals from the Lau Basin, and the Kermadec Arc. The latter sub-clade includes a paratype of *A. kermadecensis.* Neither genus *Vulcanolepas* nor *Neolepas* is monophyletic. The *Vulcanolepas/Leucolepas* from the Kermadec Arc, Lau Basin, and Mariana Arc belong to lineages that appear to have diverged earlier in history with respect to a well-supported and well-resolved clade made up by *N. zevinae/rapanuii* from the East Pacific Rise and its sister sub-clade of *V. scotiaensis* from the East Scotia Ridge and *Neolepas* sp. 1 from the Southwest and Central Indian Ridge.

Topological Hypothesis Testing

None of the phylogenetic hypotheses inferred from the Thoracica-wide concatenated Sanger-based dataset support the monophyly of barnacles from deep-sea hydrothermal vents (Fig. 2). The topological test showed that the hypothesis of monophyly was significantly less probable than the hypothesis of non-monophyly (marginal log-likelihoods -16928.21 and -16908.62 respectively). The large difference in log-likelihoods (> 5) (Kass & Raftery 1995) constitutes strong contradictory evidence against the monophyly of vent barnacles as originally suggested by Pérez-Losada *et al.* (2008).

Divergence Estimates and Biogeographic History

Tree time calibrations of the combined Sanger-based dataset produced divergence estimates slightly older under the Yule tree prior of constant speciation, when compared with the nearly identical estimates obtained under the Birth-Death prior models (Fig. 2 and supplementary electronic material). These divergence estimates are consistent with estimates from Linse *et al.* (2013). The tree obtained under the Birth-Death model had the best likelihood score; however, no significant differences were encountered

among models (log-likelihood < 1). The time to the most recent common ancestor (TMRCA) of Clade A was estimated at 68.0 million years before present (Myr BP) (95% Highest Posterior Density Interval [HPD]: 38.2-105.9) under the Birth-Death models (BD) and 79.3 Myr BP (95% HPD: 47.1-121.5) under the Yule model of constant speciation rate. The TMRCAs of Eolepadidae and the *Neolepas-Vulcanolepas-Leucolepas* sub-clade were estimated at 25.1 Myr BP (95% HPD: 12.1-43.3) and 10.5 Myr BP (95% HPD: 5.4-17.3) under BD, and 31.2 Myr BP (95% HPD: 15.4-53.7) and 13.8 Myr BP (95% HPD: 7.5-23.1) under the Yule model, respectively. Divergence between Pacific and non-Pacific *Neolepas-Vulcanolepas* eolepadids was estimated to have occurred 4.8 Myr BP (95% HPD: 2.3-8.5) and 6.4 Myr BP (95% HPD: 3.0-11.2) under BD and Yule models, respectively. The split between the East Scotia Ridge and the Indian Ocean lineages occurred 1.7 Myr BP (95% HPD: 0.4-3.8) under BD and 2.3 (95% HPD: 0.5-4.9) under Yule. The TMRCA of Clade B *Eochionelasmus* was estimated at 3.2 Myr BP (95% HPD: 1.1-6.7) under the Birth-Death model and 4.2 Myr BP (95% HPD: 1.3-8.8) under Yule. The analysis of historical biogeography suggest with high-probability that hydrothermal vent barnacles from Clade A originated in the western Pacific, and during the late Miocene –Pliocene colonized the Eastern Pacific, the Southern Ocean south of the Atlantic, and the Indian Ocean (Fig. 3).

Species Delimitation

GMYC analyses of Clade A identified a transition point between interspecific speciation–extinction processes and intraspecific coalescent processes at 0.6 Myr BP for the time-scaled combined Sanger-based phylogeny estimated with the Birth-Death model tree prior (Fig. 3). The GMYC model showed a significant ($\alpha = 0.05$) better fit to the data than the null model of uniform coalescent branching rates (likelihood ratio = 25.9, $p < 0.0001$). There were 12 distinct clusters identified, which largely corresponded to species already described or populations that were presumed to be new species. Genetic distances (*coxI* uncorrected distances) among individuals within clusters ranged between 0 and 0.9%. Genetic distances among individuals from different clusters ranged between 2 and 17.8% (except for the two *Neolepas zevinae/rapanuii* clusters whose maximum distance was 0.9%). Similarly in Clade B *Eochionelasmus* the genetic distances among individuals ranged between 0 and 0.9%.

Species Tree Estimation

The topology of the inferred Sanger-based species tree is fully congruent with the topology of the phylogenetic hypothesis obtained with the concatenated Sanger-based markers dataset, and the branch support values are mostly equal (Fig. 4). Poorly-resolved regions of the tree include the relationships among lineages of *Vulcanolepas/Leucolepas* from the Kermadec Arc, Lau Basin, Mariana Trough, and basally the positions of *Neoverruca* and *Neobrachylepas* within Clade A.

RAD Phylogenetics

RAD-seq matrices resulting from the three explored clustering thresholds (c 0.80, 0.85 and 0.90) contained similar numbers of loci and similar percentages of missing data per clustering parameter value used for the minimum taxon coverage in a given locus (approximately 15,500, 9,600, 3,800, and 600 loci, and 52 %, 44 %, 33 % and 21 % missing data, for m 4, 6, 8, and 10 respectively; see Table S8 for details). The percentages of variable sites and parsimony informative sites across matrices ranged between 6.81 - 13.18 % and 2.26 - 4.22 % respectively, being higher with smaller values of clustering thresholds and larger values of minimum taxon coverage. The tree topologies obtained from phylogenetic inferences of each matrix were identical to each other (supplementary electronic material). These topologies from RAD-seq matrices were also similar to the species tree obtained with Sanger-based markers (Fig. 4), only differing in the position of *Leucolepas*, appearing in the RAD-based trees as sister to the clade made up by *N. zevinae/rapanuii* from the East Pacific Rise, *V. scotiaensis* from the East Scotia Ridge and *Neolepas* sp. 1 from the Southwest and Central Indian Ridge. RAD-based trees topologies were highly supported with bootstrap values of 100 for all branches, except for the ones from matrices generated with a minimum taxon coverage parameter of m10. In these cases, the branches supporting the clades of *Vulcanolepas* from the Lau Basin and the Kermadec Arc, and of *Leucolepas-N. zevinae/rapanuii-V. scotiaensis- Neolepas* sp. 1 have bootstrap support values greater than 94 and 71, respectively.

DISCUSSION

Are vent barnacles monophyletic?

The inferred evolutionary history of hydrothermal vent barnacles is not consistent with the hypothesis of monophyly (single ancestry) as proposed by the smaller taxon-sampling studies of Pérez-Losada *et al.* (2008) and Linse *et al.* (2013), which included only two of the four families of vent barnacles. Our analyses of a significantly expanded dataset indicate that there are two main clades (Clade A and Clade B) (Fig. 2), thus suggesting that barnacles have colonized deep-sea hydrothermal vents at least twice in history. The results from a concurrent study by Perez-Losada *et al.* (2014) provide support to this inference by placing Clade B (*Eochionelasmus* ohtai) nested within the balanomorph barnacles, although the hypothesis of monophyly of vent barnacles was not explicitly tested in that study.

Deep-sea hydrothermal vent barnacle Clade A is the more diverse of the two, containing six of the seven genera included in this study. This clade also contains a remarkable diversity of morphologies, including asymmetric (Neoverrucidae) and symmetric (Neobrachylepadidae), pedunculate (Eolepadidae) and sessile (Neoverrucidae and Neobrachylepadidae) forms (Fig. 4) (note that neoverrucid barnacles have a

pedunculated stage during early ontogenesis (Newman & Hessler 1989)). The sister relationship of Clade A and the deep-sea pedunculate Scalpellidae (Fig. 2) (Pérez-Losada *et al.* 2008; Linse *et al.* 2013) suggests that the sessile state of the Neoverrucidae and Neobrachylepadidae is a derived state. This observation is consistent with the mounting evidence that the characters used to define higher taxonomic groups in Cirripedia need to be revised in light of multilocus molecular phylogenetic hypotheses (Pérez-Losada *et al.* 2008; Linse *et al.* 2013). A noteworthy example of this taxonomic and phylogenetic incongruence is the phylogenetic placement of *N. relica* nested in Clade A. *N. relica* is the sole living brachilepadoform species and until now was considered the most 'primitive' lineage of sessilian barnacles (Newman & Yamaguchi 1995). Clade B only contains the genus *Eochionelasmus*. Despite its morphological and phylogenetic affinities with the Balanomorpha, the phylogenetic position of *Eochionelasmus* is this study is unstable. Similarly, Perez-Losada *et al.* (2014) found low support for the branches resolving the position *Eochionelasmus ohtai* within the balanomorphs. This instability is likely caused by the long branch supporting this clade, which may indicate a rapid evolutionary rate, old divergence, or taxonomic undersampling (Fig. 2, supplementary electronic material). Further taxonomic sampling of related genera and careful review of character use for systematics should help resolve its systematics.

Deep-sea hydrothermal vent environments have been characterized as being patchy and ephemeral habitats with extreme spatial and temporal gradients of temperature, reduced chemicals, oxygen and food supply (Van Dover 2000). These conditions present significant physiological and ecological challenges to organisms and act as environmental filters that promote the evolution and distribution of species with specialized adaptations (Tunnicliffe *et al.* 2003; Fisher *et al.* 2007). The widespread persistence of vent chemosynthetic environments throughout earth's geologic history (Shock *et al.* 1995) has likely been an important factor enabling the independent colonization by multiple lineages of barnacles, as well as of other taxa, e.g., mussels (Lorion *et al.* 2013) and decapods (Yang *et al.* 2013). Clade A is nested within a predominantly deep-sea clade Linse *et al.* (2013), suggesting a colonization of hydrothermal vents at depth. The nested position within Clade A of *A. seepiophila* – the only barnacle species known to live in both cold-seep and hydrothermal vent environments – indicates a single colonization of seep environments by vent ancestors. This pattern contrasts with the stepwise colonization scenario of deep-sea chemosynthetic environments, starting in organic substrates or cold-seeps and then moving to hydrothermal vents, as suggested for other taxonomic groups, e.g., mussels (Lorion *et al.* 2013).

Historical Biogeography

The most common recent ancestor of hydrothermal-vent barnacles from Clade A likely lived in the late Mesozoic or early Cenozoic. The time to the most recent common ancestor inferred in this study is consistent with the timing inferred by Linse *et al.* (2013), but contrasts with the lower Cretaceous origin proposed by Pérez-Losada *et al.* (2008) and with classic hypotheses of antiquity of vent taxa, which proposed that hydrothermal-vent barnacles were mid-Mesozoic relict taxa (Newman 1979; Newman 1985). The discrepancy with the results from Pérez-Losada *et al.* (2008) is due to the exclusion of fossil calibration points because of uncertainty in the phylogenetic placement as described by Linse *et al.* (2013). The timing of radiation of Clade A during the Cenozoic is comparable to the estimates of origin and radiation in other chemosynthetic taxa, e.g., radiation of bresiliid shrimp 6.7-11.7 Myr BP (Shank *et al.* 1999); origin of siboglinid tubeworms ca. 60 Myr BP (Chevaldonne *et al.* 2002); radiation of chemosynthetic mussels at ca. 45 Myr BP (Lorion *et al.* 2013); radiation of kiwaid yeti crabs starting at ca. 30 Myr BP; also see reviews by Little and Vrijenhoek (2003) and Vrijenhoek (2013). A recent origin and radiation of most modern vent taxa and many other deep-sea taxa (Little & Vrijenhoek 2003; Smith & Stockley 2005; Strugnell *et al.* 2008) is consistent with the inference of a major deep-sea mass extinction event during the Cretaceous-Paleogene period boundary (Raup & Sepkoski 1982; Horne 1999; Harnik *et al.* 2012) (see Fig. 3). Several smaller-scale extinction events linked to regional Oceanic Anoxic Events, ocean acidification and temperature changes also occurred during the Cretaceous period and at the Paleocene-Eocene epoch boundary (Jacobs & Lindberg 1998; Rogers 2000; Harnik *et al.* 2012).

The most probable place of origin of the modern vent barnacle lineage from Clade A is the western Pacific, as indicated and highly supported by Bayesian ancestral state reconstruction. This is also the region where the oldest lineages and the highest diversity are found. The heterogeneity of depths in hydrothermal vent systems in the western Pacific, and the close proximity to other chemosynthetic ecosystems such as cold seeps and organic enrichments, both shallow and deep, have been suggested as important factors driving the re-colonization of vent environments and subsequent diversification (Moalic *et al.* 2011). Our analyses suggest that the most probable path of dispersal out of the western Pacific was a migration eastward during the Miocene epoch, possibly following hydrothermal vent habitats along the Pacific-Antarctic Ridge, and colonization of the eastern Pacific. The neolepadids from the East Pacific Rise have a coalescence point that is posterior to the Oligocene disruption of the Pacific-Farallon Ridge by subduction under the North American Plate, *ca.* 28.5 Myr BP (Fig. 3) (Atwater 1989), which can explain why barnacles are absent from the north-eastern Pacific vents along the Juan de Fuca Ridge. A spreading through the southern hemisphere likely followed to the East Scotia Ridge and South Sandwich Arc during the late Miocene epoch, reaching the Southwest Indian Ridge and Central Indian Ridge during

the Pliocene/Pleistocene epochs. No vent barnacle species have been found at Mid Atlantic Ridge hydrothermal vents, although the southern portion of this major mid-ocean ridge remains largely unexplored.

The proposed history of dispersal is congruent with the timing of opening of the Drake Passage during the mid Eocene epoch, *ca.* 41 Myr BP (Scher 2006), the late Eocene establishment of the eastward-flowing Antarctic Circumpolar Current (ACC), *ca.* 34 Myr BP (Scher 2006), and the mid Miocene formation of the East Scotia Rise, *ca.* 15 Myr BP (Livermore 2003) (see Fig. 3). Hydrothermal vent yeti crabs (Decapoda: Anomura: Kiwaidae) share an almost identical pattern of historical dispersal from the eastern Pacific to the East Scotia Ridge and the Southwest Indian Ridge (see Roterman *et al.* (2013) for a detailed hypothesis of vicariance in this group). A likely origin in the western or northwestern Pacific followed by migration and colonization eastward throughout the southern hemisphere during the Miocene epoch has also been inferred for other non-vent deep-sea taxa such as the octocoral *Paragorgia arborea* (Herrera *et al.* 2012), and other marine taxa such as the spiny dogfish *Squalus acanthias* (Verissimo *et al.* 2010) and the bryozoan *Membranipora membranacea* (Schwaninger 2008). These observations provide support for the biogeographic hypothesis proposed by Moalic *et al.* (2011) that the western Pacific was a centre of origin of modern vent fauna from which most taxa dispersed globally. However, our data do not support the idea of direct links between the western Pacific communities and the Indian Ocean, but rather a stepping-stone mode of dispersal in the southern hemisphere following the direction of the dominant ACC. We suggest that the geological processes and dispersal mechanisms discussed here can explain the current distribution patterns of many other marine taxa and have played an important role shaping extant deep-sea faunal diversity

The history of Clade B is not well resolved. The phylogenetic hypothesis here presented suggests that the divergence of this lineage within the Balanomorpha occurred in the Mesozoic era (Fig. 2). However this inferred antiquity is likely to be an artifact caused by taxonomic undersampling in this group. Additional data from other *Echionelasmus* populations, e.g., *E. paquensis* from the eastern Pacific, as well as from confamilial species and related groups would provide greater resolution of the evolution of Clade B.

Species Delimitation and Relationships

Inferences of species boundaries in Clade A, based on the generalized mixed Yule-coalescent method, are largely congruent with descriptions of putative morphospecies. The identified species clusters are well-constrained geographically by mid-ocean spreading ridge system and neighboring island arc basins (Figs. 3 and 4). Divergences among congeners in *Ashinkailepas* and *Neoverruca* are largely consistent with the

biogeographic boundary between the northwest and southwest Pacific, inclusive of the Mariana arc, proposed by Bachraty *et al.* (2009). Relationships among *Vulcanolepas, Leucolepas and Neolepas* species clusters remain contentious due to the non-monophyly of all three genera as defined by Buckeridge *et al.* (2013) and thus require substantial revision.

There is a lack of overlap in genetic distances for the *coxI* barcode marker within and among inferred species clusters. The maximum genetic distance within species clusters of 0.9 %, and the minimum distance among species clusters of 2 % are consistent with the proposed threshold value of *ca.* 2 % to define species boundaries through DNA barcoding in Crustacea (Hebert *et al.* 2003; Lefebure *et al.* 2006). Similarly, the maximum genetic distance among individuals of *Echionelasmus ohtai* is 0.9%. The only exception to this pattern is found in the *Neolepas zevinae/rapanuii* species cluster pair, where the maximum distance between clusters is 0.9 %. There is no phylogenetic support for this split or geographic segregation between specimens from the East Pacific Rise and southern East Pacific Rise, thus suggesting that the division of *Neolepas zevinae/rapanuii* is not indicative of species-level differentiation. The barcoding gap within and among species has been consistently found in other barnacle taxa (Tsang *et al.* 2008; Tsang *et al.* 2009; Yoshida *et al.* 2011) and in crustaceans in general (Costa *et al.* 2007; Matzen da Silva *et al.* 2011), thus our *coxI* genetic distance data provides further support to the species delimitations proposed for Clade A. The species delimitation framework developed will enable rapid species assignments as specimens from newly explored geographical regions become available.

RAD phylogenetics

Several sources of uncertainty have been associated with the use of the few traditional sequence markers available for non-model organisms (e.g., mitochondrial and ribosomal genes), including low variability, biased loci sampling, poor genomic representation, low statistical power, and inclusion of pseudogenes, among others. The effects of these are often hard to identify due to the paucity of multi-locus genome-wide comparative datasets. Such problems have been recognized and accounted for in model organisms by comparing large numbers of genomic DNA sequences from various individuals and identifying thousands of variable regions across the genome (Rokas *et al*. 2003; Clark *et al*. 2007). Recent technological and methodological developments in next-generation sequencing platforms and methodologies, such as RAD-seq, have made genomic resources increasingly accessible and available for phylogenetics in non-model organisms (Wagner *et al*. 2012; Eaton & Ree 2013; Jones *et al*. 2013; Reitzel *et al*. 2013), thus offering a great opportunity to overcome the difficulties inherent to the use of traditional approaches in many taxa.

In this study, we demonstrated that RAD-seq data provide strong support for the overall evolutionary history of vent barnacles inferred with traditional Sanger-based markers, and allow the inference of a fully resolved and supported phylogenetic tree. The small difference in topology between the species tree inferred with Sanger-based markers and the RAD-seq trees does not alter any of the conclusions regarding the biogeographical history or species delimitation of vent barnacles, but does have taxonomic implications. Further sampling and a follow-up morphological taxonomic revision would be needed to clarify the validity of the currently described genera. This study demonstrates the utility of comparative Sanger-based and RAD sequencing as a means of comparative phylogenetic inference validation in poorly known taxa such as deep-sea invertebrates.

CONCLUSIONS

Phylogenetic inferences and topology tests indicate that hydrothermal vent barnacles are not a monophyletic group. The likely timing of barnacle radiation in hydrothermal vent ecosystems was during the late Cenozoic, consistent with the timing of other specific deep-sea taxa, and correlated to the occurrence of major extinction events. Our analyses suggest that the western Pacific was the place of origin of the major hydrothermal vent barnacle lineage, followed by circumglobal colonization eastward along the southern hemisphere during the Neogene period. Inferences of species boundaries based on generalized mixed Yule-coalescent methods and DNA barcoding are largely congruent with morphological descriptions of putative species. RAD-seq data provide strong support for the overall evolutionary history inferred from Sanger-based markers and a fully resolved backbone of the vent barnacle phylogenetic tree. These results also constitute critical baseline data with which to assess potential effects of anthropogenic disturbances on deep-sea ecosystems.

AUTHOR CONTRIBUTIONS

SH and TMS conceived and designed research. SH performed research. SH analyzed data. SH, TMS and HW collected and contributed samples and reagents. SH wrote the paper with comments from TMS and HW.

DATA ACCESSIBILITY

Supplementary electronic material

Nucleotide alignments, input files, are tree files are available in Dryad DOI: doi:10.5061/dryad.7kn5k

Raw data

Raw DNA sequences are available at the U.S. National Center for Biotechnology Information (NCBI) GenBank and Sequence Read Archive (SRA accession number SRP051026).

ACKNOWLEDGEMENTS

This research was supported by the Office of Ocean Exploration and Research of the National Oceanic and Atmospheric Administration (NA09OAR4320129 to TS); the Division of Ocean Sciences of the National Science Foundation (OCE-1131620 to TS); the Division of Polar Programs of the National Science Foundation (PLR-0739675 to TS); the Astrobiology Science and Technology for Exploring Planets program of the National Aeronautics and Space Administration (NNX09AB76G to TS); and the Academic Programs Office (Ocean Ventures Fund to SH), the Ocean Exploration Institute (Fellowship support to TMS) and the Ocean Life Institute of the Woods Hole Oceanographic Institution (internal grant to TMS and SH). For enabling access to key specimens we thank K. Iizasa (U. Tokyo), Y. Suzuki (U. Tokyo), S. Nakagawa (JAMSTEC), P. Tyler (NOCS), J. Copley (NOCS), A. Rogers (Oxford), N. Roterman (Oxford), M. Clark (NIWA), A. Rowden (NIWA), K. Schnabel (NIWA), S. Mills (NIWA), J. Resing (NOAA-PMEL), R. Embley (NOAA-PMEL), A. Reysenbach (PSU), M.K. Tivey (WHOI), P. Fryer (UH), C. Langmuir (Harvard), K. Von Damm (UNH), M. Lilley (UW), the masters, crew, scientific personnel, and funding agencies of expeditions AT03-28, AT07-06, JC042, JC067, KM0417, KM0912, KOK0505, KOK0506, NT97-10, NT97-14, NT99-09, RR1211, TAN1007, TAN1104, TAN1206, TN234, TN236, YK06-13, and YK09-13. Specimens provided by the National Institute of Water and Atmospheric Research (NIWA) were collected under research programs: Kermadec Arc Minerals, funded by the New Zealand Ministry of Business, Innovation & Employment (MBIE), Auckland University, Institute of Geological and Nuclear Science (GNS), and WHOI; Ocean Survey 20/20 funded by Land Information New Zealand; Impact of resource use on vulnerable deep-sea communities (CO1X0906), funded by MBIE; and the Joint New Zealand-USA 2005 NOAA Ring of Fire Expedition, part of NIWA's Seamount Program (FRST CO1X0508). We thank A. Tarrant, A. Reitzel, and three anonymous reviewers, for providing helpful comments that improved this manuscript.

REFERENCES

Atwater T (1989) Plate tectonic history of the northeast Pacific and western North America. In: *The Eastern Pacific Ocean and Hawaii* (eds. Winterer EL, Hussong DM, Decker RW), pp. 21-72. Geological Society of America, Boulder, Colorado.

Bachraty C, Legrende P, Desbruyeres D (2009) Biogeographic relationships among hydrothermal vent faunas on a global scale. *Deep-Sea Research Part I-Oceanographic Research Papers* **56**, 1371-1378.

Baird N, Etter P, Atwood T*, et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**, e3376.

Baker ET, German CR (2004) On the global distribution of hydrothermal vent fields. *Mid-Ocean Ridges: Hydrothermal Interactions between the Lithosphere and Oceans* **148**, 245-266.

Bensasson D, Zhang D, Hartl D, Hewitt G (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology & Evolution* **16**, 314-321.

Boschen RE, Rowden AA, Clark MR, Gardner JPA (2013) Mining of deep-sea seafloor massive sulfides: A review of the deposits, their benthic communities, impacts from mining, regulatory frameworks and management strategies. *Ocean and Coastal Management* **84**, 54-67.

Brito PH, Edwards SV (2008) Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica* **135**, 439-455.

Buckeridge J (2000) *Neolepas osheai* sp. nov., a new deep-sea vent barnacle (Cirripedia: Pedunculata) from the Brothers Caldera, south-west Pacific Ocean. *New Zealand Journal of Marine and Freshwater Research* **34**, 409-418.

Buckeridge JS, Linse K, Jackson JA (2013) *Vulcanolepas scotiaensis* sp. nov., a new deep-sea scalpelliform
barnacle (Eolepadidae: Neolepadinae) from hydrothermal vents in the Scotia Sea, Antarctica. *Zootaxa* **3745**, 551-568.

Cairns S (2007) Deep-water corals: an overview with special reference to diversity and distribution of deep-water scleractinian corals. *Bulletin of Marine Science* **81**, 311-322.

Carpenter KE, Barber PH, Crandall ED*, et al.* (2011) Comparative phylogeography of the Coral Triangle and implications for marine management. *Journal of Marine Biology* **2011**, 1-14.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**, 3124-3140.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping Loci de novo from short-read sequences. *G3* **1**, 171-182.

Chevaldonne P, Jollivet D, Desbruyeres D, Lutz RA, Vrijenhoek RC (2002) Sister-species of eastern Pacific hydrothermal vent worms (Ampharetidae, Alvinellidae, Vestimentifera) provide new mitochondrial COI clock calibration. *Cahiers De Biologie Marine* **43**, 367-370.

Clark AG, Eisen MB, Smith DR*, et al.* (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203-218.

Colgan DJ, McLauchlan A, Wilson GDF*, et al.* (1998) Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolution. *Australian Journal of Zoology* **46**, 419-437.

Costa FO, deWaard JR, Boutillier J*, et al.* (2007) Biological identifications through DNA barcodes: the case of the Crustacea. *Canadian Journal of Fisheries and Aquatic Sciences* **64**, 272-295.

Cubelio SS, Tsuchida S, Watanabe S (2007) Vent associated *Munidopsis* (Decapoda : Anomura : Galatheidae) from Brothers Seamount, Kermadec Arc, Southwest Pacific, with description of one new species. *Journal of Crustacean Biology* **27**, 513-519.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**, 772.

Desbruyeres D, Alayse-Danet A-M, Ohta S, the Scientific Parties of biolauand s (1994) Deep-sea hydrothermal communities in Southwestern Pacific back-arc basins (the North Fiji and Lau Basins): Composition, microdistribution and food web. *Marine Geology* **116**, 227-242.

Drummond A, B A, S B*, et al.* (2011) Geneious v5.4. Available from http://www.geneious.com/.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**, 1969-1973.

Eaton DA (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* **30**, 1844-1849.

Eaton DAR, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: An example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology* **62**, 689-706.

Edwards SV (2008) Is a new and general theory of molecular systematics emerging? *Evolution* **63**, 1-19.

Fan Y, Wu R, Chen MH, Kuo L, Lewis PO (2011) Choosing among partition models in Bayesian phylogenetics. *Molecular Biology and Evolution* **28**, 523-532.

Fisher CR, Takai K, Le Bris N (2007) Hydrothermal vent ecosystems. *Oceanography* **20**, 14-23.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3**, 294-299.

Fujisawa T, Barraclough TG (2013) Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: A revised method and evaluation on simulated data sets. *Systematic Biology* **62**, 707-724.

Gernhard T (2008) The conditioned reconstructed process. *Journal of Theoretical Biology* **253**, 769-778.

Harnik PG, Lotze HK, Anderson SC*, et al.* (2012) Extinctions in ancient and modern seas. *Trends in Ecology & Evolution* **27**, 608-617.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences* **270**, 313-321.

Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* **27**, 570-580.

Herrera S, Munro C, Nganro N*, et al.* (2010) Biodiversity of the deep-sea benthic fauna in the Sangihe-Talaud region, Indonesia: Observations from the INDEX-SATAL 2010 expedition. *AGU Fall Meeting Abstracts* **1**, 1234.

Herrera S, Reyes-Herrera PH, Shank TM (Chapter 2) Genome-wide predictability of restriction sites across the eukaryotic tree of life.

Herrera S, Shank TM, Sanchez JA (2012) Spatial and temporal patterns of genetic variation in the widespread antitropical deep-sea coral *Paragorgia arborea*. *Molecular Ecology* **21**, 6053-6067.

Hessler RR, Lonsdale PF (1991) Biogeography of Mariana Trough hydrothermal vent communities. *Deep-Sea Research Part a-Oceanographic Research Papers* **38**, 185-199.

Hipp AL, Eaton DAR, Cavender-Bares J*, et al.* (2014) A framework phylogeny of the American Oak clade based on sequenced RAD data. *PLoS One* **9**, e93975.

Horne DJ (1999) Ocean circulation modes of the phanerozoic: implications for the antiquity of deep-sea benthonic invertebrates. *Crustaceana* **72**, 999-1018.

Jacobs DK, Lindberg DR (1998) Oxygen and evolutionary patterns in the sea: Onshore /offshore trends and recent recruitment of deep-sea faunas. *Proceedings Of The National Academy Of Sciences Of The United States Of America* **95**, 9396-9401.

Jones DS (1993) A new Neolepas (Cirripedia: Thoracica: Scalpellidae) from an abyssal hydrothermal vent, southeast Pacific. *Bulletin of Marine Science* **52**, 937.

Jones JC, Fan S, Franchini P, Schartl M, Meyer A (2013) The evolutionary history of *Xiphophorus* fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. *Molecular Ecology* **22**, 2986-3001.

Kass RE, Raftery AE (1995) Bayes factors. *Journal of the American Statistical Association* **90**, 773-795.

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**, 3059-3066.

Lefebure T, Douady CJ, Gouy M, Gilbert J (2006) Relationship between morphological taxonomy and molecular divergence within Crustacea: Proposal of a molecular threshold to help species delimitation. *Molecular Phylogenetics and Evolution* **40**, 435-447.

Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian phylogeography finds its roots. *PLoS Computational Biology* **5**, e1000520.

Linse K, Jackson JA, Fitzcharles E, Sands CJ, Buckeridge JS (2013) Phylogenetic position of Antarctic Scalpelliformes (Crustacea: Cirripedia: Thoracica). *Deep-Sea Research Part I-Oceanographic Research Papers* **73**, 99-116.

Little CTS, Vrijenhoek RC (2003) Are hydrothermal vent animals living fossils? *Trends in Ecology & Evolution* **18**, 582-588.

Livermore R (2003) Back-arc spreading and mantle flow in the East Scotia Sea. In: *Intra-oceanic subduction systems: tectonic and magmatic processes*, pp. 315-331. Geological Society of London Special Publications.

Lopez JV, Yuhki N, Masuda R, Modi W, O'Brien SJ (1994) Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *Journal of Molecular Evolution* **39**, 174-190.

Lorion J, Kiel S, Faure B*, et al.* (2013) Adaptive radiation of chemosymbiotic deep-sea mussels. *Proceedings of the Royal Society B: Biological Sciences* **280**, 20131243-20131243.

Marsh L, Copley JT, Huvenne VAI*, et al.* (2012) Microdistribution of faunal assemblages at deep-sea hydrothermal Vents in the Southern Ocean. *PLoS One* **7**, e48348.

Matzen da Silva J, Creer S, dos Santos A*, et al.* (2011) Systematic and evolutionary insights derived from mtDNA COI barcode diversity in the Decapoda (Crustacea: Malacostraca). *PLoS One* **6**, e19449.

McClain C, Mincks SL (2010) The dynamics of biogeographic ranges in the deep sea. *Proceedings of the Royal Society B: Biological Sciences* **277**, 3533–3546.

Moalic Y, Desbruyeres D, Duarte CM*, et al.* (2011) Biogeography revisited with network theory: retracing the history of hydrothermal vent communities. *Systematic Biology* **61**, 127-137.

Monaghan MT, Wild R, Elliot M*, et al.* (2009) Accelerated species inventory on madagascar using coalescent-based models of species delineation. *Systematic Biology* **58**, 298-311.

Nakamura K, Watanabe H, Miyazaki J*, et al.* (2012) Discovery of new hydrothermal activity and chemosynthetic fauna on the Central Indian Ridge at 18 degrees -20 degrees S. *PLoS One* **7**, e32965.

Newman WA (1979) A new scalpellid (Cirripedia): a Mesozoic relic living near an abyssal hydrothermal spring. *Transactions of the San Diego Society of Natural History* **19**, 153-167.

Newman WA (1985) The abyssal hydrothermal vent invertebrate fauna. A glimpse of antiquity? *Bulletin of the Biological Society of Washington* **6**, 231-242.

Newman WA, Hessler RR (1989) A new abyssal hydrothermal verrucomorphan (Cirripedia; Sessilia): the most primitive living sessile barnacle. *Transactions of the San Diego Society of Natural History* **21**, 259-273.

Newman WA, Yamaguchi T (1995) A new sessile barnacle (Cirripedia, Brachylepadomorpha) from the Lau Back-Arc Basin, Tonga; first record of a living representative since the Miocene. *Bulletin du Muséum national d'Histoire Naturelle. Section A, Zoologie, Biologie et Ecologie Animales* **17**, 221-243.

Newman WA, Yamaguchi T, Southward AJ, Segonzac M (2006) Arthropoda, Crustacea, Cirripedia. In: *Handbook of Deep-Sea Hydrothermal Vent Fauna* (eds. Desbruyeres D, Segonzac M, Bright M), pp. 356-357. Denisia.

Nylander JA, Wilgenbusch JC, Warren DL, Swofford DL (2008) AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinfomatics* **24**, 581-583.

Ohno T, Fujikura K, Hashimoto J, Fiujiwara Y, Segawa S (1996) The Hydrothermal Vent Community at the Kaikata Seamount near Ogasawara (Bonin) Islands, South Japan, pp. 221-229. (JAMSTEC) / Japan Marine Science and Technology Center (JAMSTEC).

Ohta S (1990) Deep-sea submersible survey of the hydrothermal vent community on the northeastern slope of the Iheya ridge, the Okinawa Trough. In: *The Sixth Symposium on Deep-Sea Research*, pp. 145-156. (JAMSTEC) / Japan Marine Science and Technology Center (JAMSTEC).

Pérez-Losada M, Harp M, Høeg JT*, et al.* (2008) The tempo and mode of barnacle evolution. *Molecular Phylogenetics and Evolution* **46**, 328-346.

Perez-Losada M, Hoeg JT, Simon-Blecher N*, et al.* (2014) Molecular phylogeny, systematics and morphological evolution of the acorn barnacles (Thoracica: Sessilia: Balanomorpha). *Mol Phylogenet Evol* **81**, 147-158.

Pons J, Barraclough TG, Gomez-Zurita J*, et al.* (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology* **55**, 595-609.

Rambaut A, Drummond AJ (2007) Tracer v1.4, Available from http://beast.bio.ed.ac.uk/Tracer.

Raup DM, Sepkoski JJJ (1982) Mass extinctions in the marine fossil record. *Science* **215**, 1501-1503.

Reid WDK, Sweeting CJ, Wigham BD*, et al.* (2013) Spatial differences in east scotia ridge hydrothermal vent food webs: influences of chemistry, microbiology and predation on trophodynamics. *PLoS One* **8**, e65553.

Reitzel AM, Herrera S, Layden MJ, Martindale MQ, Shank TM (2013) Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular Ecology* **22**, 2953-2970.

Rogers AD (2000) The role of the oceanic oxygen minima in generating biodiversity in the deep sea. *Deep-Sea Research Part II* **47**, 119-148.

Rogers AD, Tyler PA, Connelly DP*, et al.* (2012) The discovery of new deep-sea hydrothermal vent communities in the southern ocean and implications for biogeography. *PLoS Biology* **10**, e1001234.

Rokas A, Williams BL, King N, Carroll SB (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**, 798-804.

Ronquist F, Teslenko M, van der Mark P*, et al.* (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**, 539-542.

Roterman CN, Copley JT, Linse KT, Tyler PA, Rogers AD (2013) The biogeography of the yeti crabs (Kiwaidae) with notes on the phylogeny of the Chirostyloidea (Decapoda: Anomura). *Proceedings of the Royal Society B: Biological Sciences* **280**, 20130718-20130718.

Scher HD (2006) Timing and climatic consequences of the opening of Drake Passage. *Science* **312**, 428-430.

Schwaninger HR (2008) Global mitochondrial DNA phylogeography and biogeographic history of the antitropically and longitudinally disjunct marine bryozoan *Membranipora membranacea* L. (Cheilostomata): Another cryptic marine sibling species complex? *Molecular Phylogenetics and Evolution* **49**, 893-908.

Shank T, Herrera S, Bors E*, et al.* (2010) Hydrothermal vents and organic falls in the heart of the Coral Triangle: chemosynthetic communities discovered via relepresence in the Sangihe-Talaud region, Northern Sulawesi, Indonesia. *AGU Fall Meeting Abstracts* **1**, 04.

Shank TM, Black MB, Halanych KM, Lutz RA, Vrijenhoek RC (1999) Miocene radiation of deep-sea hydrothermal vent shrimp (Caridea: Bresiliidae): evidence from mitochondrial cytochrome oxidase subunit I. *Molecular Phylogenetics and Evolution* **13**, 244-254.

Shock EL, Mccollom T, Schulte MD (1995) Geochemical Constraints on Chemolithoautotrophic Reactions in Hydrothermal Systems. *Origins of Life and Evolution of the Biosphere* **25**, 141-159.

Smith AB, Stockley B (2005) The geological history of deep-sea colonization by echinoids: roles of surface productivity and deep-water ventilation. *Proceedings of the Royal Society B: Biological Sciences* **272**, 865-869.

Southward A (2005) Systematics and ecology of a new species of stalked barnacle (Cirripedia: Thoracica: Scalpellomorpha: Eolepadidae: Neolepadini) from the Pacific-Antarctic Ridge at 38° S. *Marine Biodiversity* **35**, 147-156.

Southward AJ, Newman WA (1998) Ectosymbiosis between filamentous sulphur bacteria and a stalked barnacle (Scalpellomorpha, Neolepadinae) from the Lau Back Arc Basin, Tonga. *Cahiers de Biologie Marine* **39**, 259-262.

Stadler T (2009) On incomplete sampling under birth-death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology* **261**, 58-66.

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690.

Strugnell JM, Rogers AD, Prodohl PA, Collins M, Allcock AL (2008) The thermohaline expressway: the Southern Ocean as a centre of origin for deep-sea octopuses. *Cladistics* **24**, 853-860.

Tao C, Lin J, Guo S*, et al.* (2011) First active hydrothermal vents on an ultraslow-spreading center: Southwest Indian Ridge. *Geology* **40**, 47-50.

Thurber AR, Sweetman AK, Narayanaswamy BE*, et al.* (2014) Ecosystem function and services provided by the deep sea. *Biogeosciences* **11**, 3941-3963.

Tsang L, Chan B, Wu T*, et al.* (2008) Population differentiation in the barnacle *Chthamalus malayensis*: postglacial colonization and recent connectivity across the Pacific and Indian Oceans. *Marine Ecology Progress Series* **364**, 107-118.

Tsang LM, Chan BKK, Shih F-L, Chu KH, Allen Chen C (2009) Host-associated speciation in the coral barnacle *Wanella milleporae* (Cirripedia: Pyrgomatidae) inhabiting the *Millepora* coral. *Molecular Ecology* **18**, 1463-1475.

Tufar W (1990) Modern hydrothermal activity, formation of complex massive sulfide deposits and associated vent communities in the Manus back-arc basin (bismarck Sea, Papua New Guinea). *Mitteilungen der Oesterreichischen Geologischen Gesellschaft* **82**, 183-210.

Tunnicliffe V, Juniper SK, Sibuet M (2003) Reducing environments of the deep-sea floor. In: *Ecosystems of the World: The Deep Sea* (ed. Tyler PA), pp. 81-110, Amsterdam.

Tunnicliffe V, Southward AJ (2004) Growth and breeding of a primitive stalked barnacle *Leucolepas longa* (Cirripedia: Scalpellomorpha: Eolepadidae: Neolepadinae) inhabiting a volcanic seamount off Papua New Guinea. *Journal of the Marine Biological Association of the UK* **84**, 121-132.

Van Dover CL (2000) *The Ecology of Deep-Sea Hydrothermal Vents* Princeton University Press, Princeton, N.J.

Van Dover CL (2002) Trophic relationships among invertebrates at the Kairei hydrothermal vent field (Central Indian Ridge). *Marine Biology* **141**, 761-772.

Van Dover CL (2010) Mining seafloor massive sulphides and biodiversity: what is at risk? *ICES Journal of Marine Science* **68**, 341-348.

Van Dover CL, Humphris SE, Fornari D*, et al.* (2001) Biogeography and ecological setting of Indian Ocean hydrothermal vents. *Science* **294**, 818-823.

Van Dover CL, Smith CR, Ardron J*, et al.* (2012) Designating networks of chemosynthetic ecosystem reserves in the deep sea. *Marine Policy* **36**, 378-381.

Verissimo A, McDowell JR, Graves JE (2010) Global population structure of the spiny dogfish *Squalus acanthias*, a temperate shark with an antitropical distribution. *Molecular Ecology* **19**, 1651-1662.

Vrijenhoek R (2013) On the instability and evolutionary age of deep-sea chemosynthetic communities. *Deep-Sea Research Part II* **92**, 189–200.

Wagner CE, Keller I, Wittwer S*, et al.* (2012) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology* **22**, 787-798.

Watanabe H, Tsuchida S, Fujikura K*, et al.* (2005) Population history associated with hydrothermal vent activity inferred from genetic structure of neoverrucid barnacles around Japan. *Marine Ecology-Progress Series* **288**, 233-240.

Whiting MF (2002) Mecoptera is paraphyletic: multiple genes and phylogeny of Mecoptera and Siphonaptera. *Zoologica Scripta* **31**, 93-104.

Xia X, Xie Z, Salemi M, Chen L, Wang Y (2003) An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution* **26**, 1-7.

Xia XH (2013) DAMBE5: A comprehensive software package for data analysis in molecular biology and evolution. *Molecular Biology and Evolution* **30**, 1720-1728.
Xie WG, Lewis PO, Fan Y, Kuo L, Chen MH (2011) Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology* **60**, 150-160.
Yang JS, Lu B, Chen DF*, et al.* (2013) When did decapods invade hydrothermal vents? Clues from the western Pacific and Indian Oceans. *Molecular biology and evolution* **30**, 305-309.
Yoshida R, Osawa M, Hirose M, Hirose E (2011) A new fenus and two new species of Peltogastridae (Crustacea: Cirripedia: Rhizocephala) parasitizing Hermit Crabs from Okinawa Island (Ryukyu Archipelago, Japan), and their DNA-barcodes. *Zoological Science* **28**, 853-862.
Yule GA (1925) A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis. *Philosophical Transactions of the Royal Society of London B Biological Sciences* **213**, 21-87.

SUPPORTING INFORMATION

Table S1. Collection and sequence information for the specimens used in this study.

Species	Region	Locality	Cruise	Collection Platform	Collection Event	Date of Collection YYYYMMDD	Depth (m)	Lat.	Long.	ID sample
Ashinkailepas seepiophila	Izu-Ogasawara Arc	Myojin Knoll	NT97-10	HOV Shinkai 2000	Dive 952	19970622	1268	32.100	139.875	AsOg1
Ashinkailepas seepiophila	Izu-Ogasawara Arc	Myojin Knoll	NT97-10	HOV Shinkai 2000	Dive 952	19970622	1268	32.100	139.875	AsOg2
Ashinkailepas seepiophila	Izu-Ogasawara Arc	Myojin Knoll	NT97-10	HOV Shinkai 2000	Dive 952	19970622	1268	32.100	139.875	AsOg3
Ashinkailepas seepiophila	Izu-Ogasawara Arc	Myojin Knoll	NT97-10	HOV Shinkai 2000	Dive 952	19970622	1268	32.100	139.875	AsOg4
Ashinkailepas seepiophila	Okinawa Trough	Iheya Ridge	NT97-14	HOV Shinkai 2000	Dive 977	19970924	1396	27.545	126.972	AsOk1
Ashinkailepas seepiophila	Okinawa Trough	Iheya Ridge	NT97-14	HOV Shinkai 2000	Dive 977	19970924	1396	27.545	126.972	AsOK2
Ashinkailepas seepiophila	Okinawa Trough	Iheya Ridge	NT97-14	HOV Shinkai 2000	Dive 977	19970924	1396	27.545	126.972	AsOK3
Ashinkailepas_kermadecensis	Kermader Arc	Wright Seamount	KOK0505	HOV Pisces V	Dive 621	20050418	1165	-31.861	-179.188	18008
Ashinkailepas_kermadecensis	Lau Basin	Niua Seamount	RR1211	ROV Quest 4000	Dive 330	20120922	723	-15.081	-173.553	AsNi1
Ashinkailepas_kermadecensis	Tonga Arc	Niua Seamount	RR1211	ROV Quest 4000	Dive 330	20120922	723	-15.081	-173.553	AsNiN1
Ashinkailepas_kermadecensis	Tonga Arc	Niua Seamount	RR1211	ROV Quest 4000	Dive 330	20120922	723	-15.081	-173.553	AsNiN2
Ashinkailepas_kermadecensis	Tonga Arc	Niua Seamount	RR1211	ROV Quest 4000	Dive 330	20120922	723	-15.081	-173.553	AsNiN3
Ashinkailepas_kermadecensis	Tonga Arc	Niua Seamount	RR1211	ROV Quest 4000	Dive 330	20120922	723	-15.081	-173.553	AsNiN4
Eochionelasmus ohtai	Lau Basin	ELSC	KM0417	Scripps-type dredge	Dredge 52	20041009	2640	-20.050	-176.134	bar11
Eochionelasmus ohtai	Lau Basin	ELSC	KM0417	Scripps-type dredge	Dredge 52	20041009	2640	-20.050	-176.134	bar13
Eochionelasmus ohtai	Lau Basin	ELSC	KM0417	Scripps-type dredge	Dredge 52	20041009	2640	-20.050	-176.134	bar14
Eochionelasmus ohtai	Lau Basin	ELSC	TN236	ROV Jason 2	Dive 444	20090627	2232	-20.682	-176.183	bar20
Eochionelasmus ohtai	Lau Basin	ELSC	TN236	ROV Jason 2	Dive 444	20090627	2232	-20.682	-176.183	bar21
Eochionelasmus ohtai	Lau Basin	Fonualei South Volcano	RR1211	ROV Quest 4000	Dive 323	20120913	956	-17.542	-174.576	EoFoS1
Eochionelasmus ohtai	Lau Basin	Fonualei South Volcano	RR1211	ROV Quest 4000	Dive 323	20120913	956	-17.542	-174.576	EoFoS2
Eochionelasmus ohtai	Lau Basin	Fonualei South Volcano	RR1211	ROV Quest 4000	Dive 323	20120913	956	-17.542	-174.576	EoFoS3
Eochionelasmus ohtai	Lau Basin	NELSC	RR1211	ROV Quest 4000	Dive 323	20120913	956	-17.542	-174.576	EoFoS4
Eochionelasmus ohtai	Lau Basin	NELSC	TN234	ROV Jason 2	Dive 415	20090508	1617	-15.383	-174.245	EoNE1
Eochionelasmus ohtai	Lau Basin	NELSC	TN234	ROV Jason 2	Dive 415	20090508	1617	-15.383	-174.245	EoNE3
Eochionelasmus ohtai	Lau Basin	NELSC	TN234	ROV Jason 2	Dive 415	20090508	1617	-15.383	-174.245	EoNE4
Eochionelasmus ohtai	Tonga Arc	Mata Ua Volcano	RR1211	ROV Quest 4000	Dive 328	20120920	2391	-15.017	-173.788	EoMaU1
Eochionelasmus ohtai	Tonga Arc	Mata Ua Volcano	RR1211	ROV Quest 4000	Dive 328	20120920	2391	-15.017	-173.788	EoMaU2
Eochionelasmus ohtai	Tonga Arc	Mata Ua Volcano	RR1211	ROV Quest 4000	Dive 328	20120920	2391	-15.017	-173.788	EoMaU3
Eochionelasmus ohtai	Tonga Arc	Mata Ua Volcano	RR1211	ROV Quest 4000	Dive 328	20120920	2391	-15.017	-173.788	EoMaU4
Eochionelasmus ohtai manusensis	Manus Basin	Vienna Woods	YK06-13	HOV Shinkai 6500	Dive 980	20060918	2477	-3.163	150.279	EoMa1
Eochionelasmus ohtai manusensis	Manus Basin	Vienna Woods	YK06-13	HOV Shinkai 6500	Dive 980	20060918	2477	-3.163	150.279	EoMa2
Eochionelasmus ohtai manusensis	Manus Basin	Vienna Woods	YK06-13	HOV Shinkai 6500	Dive 980	20060918	2477	-3.163	150.279	EoMa3
Eochionelasmus ohtai manusensis	Manus Basin	Vienna Woods	YK06-13	HOV Shinkai 6500	Dive 980	20060918	2477	-3.163	150.279	EoMa4
Leucolepas longa	Mariana Arc	TOTO Caldera	KM0912	HROV Nereus	Dive 015	20090604	2949	12.711	143.543	bar01
Leucolepas longa	Mariana Arc	TOTO Caldera	KM0912	HROV Nereus	Dive 015	20090604	2949	12.711	143.543	bar02
Leucolepas longa	Mariana Arc	TOTO Caldera	KM0912	HROV Nereus	Dive 015	20090604	2949	12.711	143.543	VuTO1

Leucolepas longa	Mariana Arc	TOTO Caldera	KM0912	HROV Nereus	Dive 015	20090604	2949	12.711	143.543	VuTO2
Leucolepas longa	Mariana Arc	TOTO Caldera	KM0912	HROV Nereus	Dive 015	20090604	2949	12.711	143.543	VuTO3
Leucolepas longa	Mariana Arc	TOTO Caldera	KM0912	HROV Nereus	Dive 015	20090604	2949	12.711	143.543	VuTO4
Neobrachylepas relica	Lau Basin	NELSC	TN234	ROV Jason 2	Dive 415	20090508	1618	-15.383	-174.245	bar03
Neobrachylepas relica	Lau Basin	NELSC	TN234	ROV Jason 2	Dive 415	20090508	1618	-15.383	-174.245	bar04
Neobrachylepas relica	Lau Basin	NELSC	TN234	ROV Jason 2	Dive 415	20090508	1618	-15.383	-174.245	bar05
Neobrachylepas relica	Lau Basin	NELSC	TN234	ROV Jason 2	Dive 415	20090508	1618	-15.383	-174.245	bar06
Neobrachylepas relica	Lau Basin	NELSC	TN234	ROV Jason 2	Dive 415	20090508	1618	-15.383	-174.245	bar07
Neobrachylepas relica	Lau Basin	NELSC	TN234	ROV Jason 2	Dive 415	20090508	1618	-15.383	-174.245	bar08
Neobrachylepas relica	Lau Basin	NELSC	TN234	ROV Jason 2	Dive 415	20090508	1618	-15.383	-174.245	EoNE2
Neobrachylepas relica	Lau Basin	NELSC	TN234	ROV Jason 2	Dive 415	20090508	1618	-15.383	-174.245	NeNE1
Neobrachylepas relica	Lau Basin	NELSC	TN234	ROV Jason 2	Dive 415	20090508	1618	-15.383	-174.245	NeNE2
Neobrachylepas relica	Lau Basin	NELSC	TN234	ROV Jason 2	Dive 415	20090508	1618	-15.383	-174.245	NeNE5
Neobrachylepas relica	Lau Basin	NELSC	TN234	ROV Jason 2	Dive 415	20090508	1618	-15.383	-174.245	NeNE6
Neobrachylepas relica	Lau Basin	NELSC	TN234	ROV Jason 2	Dive 415	20090508	1618	-15.383	-174.245	NeNE7
Neobrachylepas relica	Lau Basin	NELSC	TN234	ROV Jason 2	Dive 415	20090508	1618	-15.383	-174.245	NeNE8
Neobrachylepas relica	Lau Basin	NELSC	TN234	ROV Jason 2	Dive 415	20090508	1618	-15.383	-174.245	NerNE3
Neobrachylepas relica	Lau Basin	NELSC	TN234	ROV Jason 2	Dive 415	20090508	1618	-15.383	-174.245	NerNE4
Neolepas sp. 1	CIR	Kairei Field	YK09-13	HOV Shinkai 6500	Dive 1175	20091113	2422	-25.320	70.040	NeIn1
Neolepas sp. 1	CIR	Kairei Field	YK09-13	HOV Shinkai 6500	Dive 1175	20091113	2422	-25.320	70.040	NeIn2
Neolepas sp. 1	CIR	Kairei Field	YK09-13	HOV Shinkai 6500	Dive 1175	20091113	2422	-25.320	70.040	NeIn3
Neolepas sp. 1	CIR	Kairei Field	YK09-13	HOV Shinkai 6500	Dive 1175	20091113	2422	-25.320	70.040	NeIn4
Neolepas sp. 1	SWIR	Dragon Field	JC067	ROV Kiel 6000	Dive 1	20111127	2770	-37.784	49.649	barJC6731B11
Neolepas sp. 1	SWIR	Dragon Field	JC067	ROV Kiel 6000	Dive 1	20111127	2770	-37.784	49.649	barJC6731B12
Neolepas sp. 1	SWIR	Dragon Field	JC067	ROV Kiel 6000	Dive 1	20111127	2770	-37.784	49.649	barJC6731B13
Neolepas zevinae/rapanuii	East Pacific Rise	9 50'N	AT07-06	HOV Alvin	Dive 3754	20020116	2499	9.827	-104.292	bar22
Neolepas zevinae/rapanuii	East Pacific Rise	9 50'N	AT07-06	HOV Alvin	Dive 3754	20020118	2490	9.848	-104.289	bar23
Neolepas zevinae/rapanuii	SEPR	17 S	AT03-28	HOV Alvin	Dive 3294	19981025	2573	-17.418	-113.204	bar15
Neolepas zevinae/rapanuii	SEPR	17 S	AT03-28	HOV Alvin	Dive 3294	19981025	2573	-17.418	-113.204	SEPR1
Neolepas zevinae/rapanuii	SEPR	17 S	AT03-28	HOV Alvin	Dive 3294	19981025	2573	-17.418	-113.204	SEPR2
Neolepas zevinae/rapanuii	SEPR	17 S	AT03-28	HOV Alvin	Dive 3294	19981025	2573	-17.418	-113.204	SEPR3
Neolepas zevinae/rapanuii	SEPR	17 S	AT03-28	HOV Alvin	Dive 3294	19981025	2573	-17.418	-113.204	SEPR4
Neoverruca sp. 1	Izu-Ogasawara Arc	Myojin Knoll	NT99-09	HOV Shinkai 2000	Dive 1112	19990629	1340	32.105	139.867	NeOg1
Neoverruca sp. 1	Izu-Ogasawara Arc	Myojin Knoll	NT99-09	HOV Shinkai 2000	Dive 1112	19990629	1340	32.105	139.867	NeOg2
Neoverruca sp. 1	Izu-Ogasawara Arc	Myojin Knoll	NT99-09	HOV Shinkai 2000	Dive 1112	19990629	1340	32.105	139.867	NeOg3
Neoverruca sp. 1	Izu-Ogasawara Arc	Myojin Knoll	NT99-09	HOV Shinkai 2000	Dive 1112	19990629	1340	32.105	139.867	NeOg4
Vulcanolepas osheai	Kermader Arc	Brothers Seamount	TAN1007	Epibenthic Sledge	Station 080	20100604	1342	-34.879	179.070	595B2
Vulcanolepas osheai	Kermader Arc	Brothers Seamount	TAN1007	Epibenthic Sledge	Station 080	20100604	1342	-34.879	179.070	595G
Vulcanolepas osheai	Kermader Arc	Brothers Seamount	TAN1007	Epibenthic Sledge	Station 079	20100604	1437	-34.878	179.071	bar592C
Vulcanolepas osheai	Kermader Arc	Brothers Seamount	TAN1007	Epibenthic Sledge	Station 080	20100604	1342	-34.879	179.070	bar595D4
Vulcanolepas osheai	Kermader Arc	Clark Seamount	KOK0506	HOV Pisces V	Dive 623	20050428	884	-36.447	177.839	337451
Vulcanolepas osheai	Kermader Arc	Clark Seamount	KOK0506	HOV Pisces V	Dive 623	20050428	884	-36.447	177.839	337452

Vulcanolepas osheai	Kermader Arc	Clark Seamount	KOK0506	HOV Pisces V	Dive 623	20050428	884	-36.447	177.839	337453
Vulcanolepas osheai	Kermader Arc	Clark Seamount	KOK0506	HOV Pisces V	Dive 623	20050428	884	-36.447	177.839	337454
Vulcanolepas osheai	Kermader Arc	Healy Seamount	TAN1104	Epibenthic Sledge	Station 073	20110313	1255	-35.014	178.980	72638_22
Vulcanolepas osheai	Kermader Arc	Tangaroa Seamount	TAN1206	Epibenthic Sledge	Station 017	20120416	682	-36.325	178.031	82121_15
Vulcanolepas scotiaensis	East Scotia Rise	E2	JC042	ROV Isis	Dive 133	20100123	2700	-56.060	-30.330	16640
Vulcanolepas scotiaensis	East Scotia Rise	E2	JC042	ROV Isis	Dive 133	20100123	2700	-56.060	-30.330	16641
Vulcanolepas scotiaensis	East Scotia Rise	E2	JC042	ROV Isis	Dive 133	20100123	2700	-56.060	-30.330	16642
Vulcanolepas scotiaensis	East Scotia Rise	E2	JC042	ROV Isis	Dive 133	20100123	2700	-56.060	-30.330	16643
Vulcanolepas scotiaensis	East Scotia Rise	E9	JC042	ROV Isis	Dive 141	20100130	2400	-60.050	-29.930	46923
Vulcanolepas scotiaensis	East Scotia Rise	E9	JC042	ROV Isis	Dive 141	20100130	2400	-60.050	-29.930	46924
Vulcanolepas scotiaensis	East Scotia Rise	E9	JC042	ROV Isis	Dive 141	20100130	2400	-60.050	-29.930	46925
Vulcanolepas scotiaensis	East Scotia Rise	E9	JC042	ROV Isis	Dive 141	20100130	2400	-60.050	-29.930	46926
Vulcanolepas sp. 1	Lau Basin	NELSC	TN234	ROV Jason 2	Dive 415	20090508	1617	-15.383	-174.245	bar09
Vulcanolepas sp. 1	Tonga Arc	Mata Ua Volcano	RR1211	ROV Quest 4000	Dive 328	20120920	2391	-15.017	-173.788	VuMaU1
Vulcanolepas sp. 1	Tonga Arc	Mata Ua Volcano	RR1211	ROV Quest 4000	Dive 328	20120920	2391	-15.017	-173.788	VuMaU2
Vulcanolepas sp. 1	Tonga Arc	Mata Ua Volcano	RR1211	ROV Quest 4000	Dive 328	20120920	2391	-15.017	-173.788	VuMaU3
Vulcanolepas sp. 1	Tonga Arc	Mata Ua Volcano	RR1211	ROV Quest 4000	Dive 328	20120920	2391	-15.017	-173.788	VuMaU4

Abbreviations: Central Indian Ridge (CIR), Eastern Lau Spreading Center (ELSC), North-East Lau Spreading Center (NELSC), southern East Pacific Rise (SEPR), Southwest Indian Ridge (SWIR)

Table S2. Accession numbers for sequences from the Superorder Thoracica retrieved from GenBank

cox1

Taxon ID	GenBank Accession		
Conchoderma hunteri	KC138462	*Amphibalanus variegatus*	KC138446
Conchoderma hunteri	KC138463	*Amphibalanus variegatus*	KC138447
Conchoderma virgatum	KC138464	*Galkinia indica*	JQ946272
Lepas testudinata	KC138477	*Galkinia indica*	JQ946215
Lepas testudinata	KC138478	*Galkinia indica*	JQ946238
Paralepas sp HNC2013	KC138502	*Galkinia indica*	JQ946273
Lepas anserifera	KC138474	*Galkinia equus*	JQ946276
Heteralepas japonica	FJ694788	*Galkinia equus*	JQ946251
Heteralepas sp LMT2008	EU884170	*Galkinia tabulatus*	JQ946255
Heteralepas sp LMT2008	EU884172	*Galkinia altiapiculus*	JQ946270
Heteralepas japonica	EU884154	*Galkinia altiapiculus*	JQ946287
Heteralepas japonica	EU884153	*Galkinia altiapiculus*	JQ946271
Heteralepas japonica	EU884169	*Galkinia altiapiculus*	JQ946286
Paralepas sp LMT2008	EU884174	*Galkinia depressa*	JQ946289
Heteralepas sp LMT2008	EU884171	*Galkinia depressa*	JQ946285
Pseudoctomeris sulcata	KC138503	*Galkinia depressa*	JQ946288
Pseudoctomeris sulcata	KC138504	*Armatobalanus allium*	KC138450
Chthamalus sp HNC2013	KC138465	*Cantellius hoegi*	FJ379314
Chthamalus sp HNC2013	KC138466	*Cantellius hoegi*	KC138453
Hexechamaesipho pilsbryi	KC138470	*Cantellius hoegi*	FJ379315
Hexechamaesipho pilsbryi	KC138471	*Cantellius hoegi*	KC138454
Galkinia depressa	JQ946262	*Cantellius sextus*	FJ379311
Megabalanus ajax	KC138480	*Cantellius sextus*	KC138456
Megabalanus ajax	KC138479	*Cantellius sextus*	FJ379312
Megabalanus occator	KC138483	*Cantellius sextus*	KC138457
Megabalanus occator	KC138484	*Nobia grandis*	FJ379318
Megabalanus zebra	KC138491	*Nobia grandis*	KC138496
Megabalanus zebra	KC138492	*Nobia grandis*	KC138497
Megabalanus tintinnabulum	KC138487	*Balanus trigonus*	KC138451
Megabalanus tintinnabulum	KC138488	*Balanus trigonus*	KC138452
Megabalanus volcano	KC138489	*Wanella milleporae*	EF565204
Megabalanus volcano	KC138490	*Wanella milleporae*	KC138514
Megabalanus rosa	KC138485	*Wanella milleporae*	KC138515
Megabalanus rosa	KC138486	*Amphibalanus amphitrite*	KC138445
Megabalanus coccopoma	KC138481	*Fistulobalanus albicostatus*	KC138468
Megabalanus coccopoma	KC138482	*Fistulobalanus albicostatus*	KC138469
Amphibalanus zhujiangensis	KC138448	*Membranobalanus longirostrum*	KC138493
Darwiniella conjugatum	JQ988964	*Cantellius pallidus*	FJ379317
Nobia conjugatum	EF569557	*Wanella milleporae*	EF565206
Darwiniella conjugatum	JQ988945	*Tetraclitella karandei*	KC138510
Darwiniella conjugatum	JQ988946	*Tetraclitella karandei*	KC138511
Darwiniella conjugatum	JQ989033	*Tetraclitella chinensis*	KC138506
Nobia conjugatum	EF569556	*Yamaguchiella sp HNC2013*	KC138512
Cionophorus soongi	JQ946277	*Yamaguchiella sp HNC2013*	KC138513
Cionophorus soongi	JQ946278	*Newmanella sp HNC2013*	KC138494
Cionophorus soongi	JQ946279	*Newmanella sp HNC2013*	KC138495
Darwiniella angularis	JQ988974	*Chelonibia patula*	JF823664
Darwiniella angularis	JQ988978	*Chelonibia patula*	JF823661
Darwiniella angularis	JQ988980	*Chelonibia testudinaria*	KC138460
Hiroa stubbingsi	JQ946244	*Chelonibia patula*	JF823663
Hiroa stubbingsi	JQ946246	*Chelonibia patula*	JF823662
Hiroa stubbingsi	JQ946237	*Chelonibia testudinaria*	KC138461
Hiroa stubbingsi	JQ946245	*Chelonibia testudinaria*	JF823739

Chelonibia testudinaria	JF823740	*Tetraclita formosana*	DQ647761
Tetraclita kuroshioensis	KC138505	*Tetraclita japonica*	DQ647704
Tetraclitella chinensis	KC138507	*Tetraclita japonica*	DQ647707
Darwiniella angularis	JQ988981	*Tetraclita formosana*	DQ647764
Capitulum mitella	KC138458	*Tetraclita japonica*	DQ647706
Capitulum mitella	KC138459	*Tetraclita formosana*	DQ647763
Ibla cumingi	KC138472	*Tetraclita japonica*	DQ647705
Ibla cumingi	KC138473	*Tetraclita formosana*	DQ647762
Galkinia equus	JQ946226	*Tetraclita singaporensis*	EF035167
Tetraclitella divisa	KC138508	*Tesseropora rosea*	FJ516076
Tetraclitella divisa	KC138509	*Tesseropora rosea*	FJ516077
Chthamalus sp 1 JDZ2005	AY823025	*Tesseropora rosea*	FJ516074
Tesseropora wireni pacifica	AY823032	*Tesseropora rosea*	FJ516075
Amphibalanus improvisus	FJ845840	*Tetraclita rubescens*	GU381926
Trevathana mizrachae	FJ620805	*Catomerus polymerus*	FJ516172
Trevathana mizrachae	FJ620802	*Catomerus polymerus*	FJ516173
Trevathana mizrachae	FJ620803	*Catomerus polymerus*	FJ516175
Trevathana mizrachae	FJ620804	*Catomerus polymerus*	FJ516174
Trevathana margaretae	FJ620810	*Tetraclita rubescens*	GU381927
Trevathana margaretae	FJ620811	*Chthamalus cf challengeri JDZ2005*	AY823019
Trevathana margaretae	FJ620812	*Tetrachthamalus sinensis*	JQ755178
Trevathana dentata	FJ620820	*Chinochthamalus scutelliformis*	JQ755182
Trevathana dentata	FJ620821	*Chthamalus fragilis*	JQ755179
Trevathana dentata	FJ620822	*Chthamalus bisinuatus*	FJ845849
Trevathana jensi	FJ620823	*Conopea sp A DCS2011*	JQ966291
Trevathana jensi	FJ620824	*Tetraclita rubescens*	GU381925
Trevathana jensi	FJ620825	*Chthamalus moro*	EU304448
Trevathana jensi	FJ620826	*Tetraclitella purpurascens*	FJ516126
Trevathana sarae	FJ620798	*Tetraclitella purpurascens*	FJ516110
Trevathana sarae	FJ620800	*Tetraclitella purpurascens*	FJ516127
Savignium crenatum	FJ620792	*Tetraclitella purpurascens*	FJ516125
Savignium crenatum	FJ620789	*Hexechamaesipho pilsbryi*	KC896275
Savignium crenatum	FJ620791	*Pollicipes pollicipes*	EF462950
Neotrevathana elongatum	FJ620831	*Pollicipes pollicipes*	EF462951
Trevathana sarae	FJ620830	*Pseudoctomeris sulcata*	JX083865
Trevathana sarae	FJ620799	*Chamaesipho tasmanica*	JX083867
Savignium crenatum	FJ620790	*Capitulum mitella*	JX502998
Trevathana margaretae	FJ620813	*Chthamalus angustitergum*	FJ845832
Semibalanus balanoides	FJ845815	*Chthamalus angustitergum*	FJ845836
Tetraclita rubescens	GU381928	*Chthamalus angustitergum*	FJ845839
Lepas anatifera	GU993589	*Chthamalus angustitergum*	FJ845833
Lepas anatifera	GU993590	*Chthamalus bisinuatus*	FJ845850
Lepas anatifera	GU993591	*Chthamalus bisinuatus*	FJ845845
Lepas pectinata	GU993645	*Chthamalus bisinuatus*	FJ845846
Lepas pectinata	GU993644	*Microeuraphia depressa*	HQ224880
Lepas pectinata	GU993650	*Microeuraphia depressa*	EF095159
Lepas anserifera	GU993630	*Microeuraphia depressa*	EF095160
Lepas anserifera	GU993629	*Chthamalus montagui*	FJ858061
Lepas australis	GU993638	*Chthamalus montagui*	FJ858066
Lepas australis	GU993639	*Chthamalus montagui*	FJ858062
Lepas australis	GU993641	*Chthamalus montagui*	FJ858060
Lepas australis	GU993640	*Chthamalus stellatus*	EU699247
Lepas anatifera	GU993588	*Chthamalus stellatus*	EU699241
Tetrachthamalus oblitteratus	AY430813	*Microeuraphia rhizophorae*	FJ845864
Chamaesipho sp BOLDAAW6872	DQ889089	*Microeuraphia rhizophorae*	FJ845865
Tetraclita singaporensis	EF035166	*Microeuraphia rhizophorae*	FJ845866
Tetraclita singaporensis	EF035164	*Euraphia eastropacensis*	FJ845851
Tetraclita singaporensis	EF035165	*Euraphia eastropacensis*	FJ845852

Euraphia eastropacensis	FJ845858
Chthamalus neglectus	FJ858077
Chthamalus neglectus	FJ858079
Chthamalus neglectus	FJ858080
Chthamalus neglectus	FJ858078
Chthamalus malayensis	FJ845828
Chthamalus malayensis	FJ845830
Chthamalus malayensis	FJ845831
Chthamalus malayensis	FJ845829
Chthamalus dentatus	FJ858084
Chthamalus dentatus	FJ858086
Chthamalus dentatus	FJ858087
Chthamalus dentatus	FJ858088
Chthamalus antennatus	JX083870
Euraphia eastropacensis	FJ845857
Chthamalus cortezianus	AF234810
Chthamalus cortezianus	AF234811
Chthamalus cortezianus	AF234812
Microeuraphia sp 2 MPL2012	JX083873
Euraphia sp 2 MPL2012	JX083872
Chthamalus challengeri	FJ858069
Chthamalus challengeri	FJ858073
Chthamalus challengeri	FJ858074
Chthamalus challengeri	FJ858075
Hexechamaesipho pilsbryi	JX083868
Caudoeuraphia caudata	JX083871
Megatrema anglicum	FJ713101
Nesochthamalus intertextus	JX083869
Megabalanus rosa	JX503004
Megabalanus rosa	JX503005
Balanus glandula	EF694592
Balanus glandula	EF694596
Balanus glandula	EF694594
Balanus glandula	EF694595
Darwiniella sp HNC2013	KC138467
Conopea galeata	JQ966287
Conopea galeata	JQ966288
Chamaesipho columna	JX083866
Semibalanus cariosus	GQ902335
Semibalanus cariosus	GQ902333
Semibalanus cariosus	GQ902332
Semibalanus cariosus	GQ902334
Semibalanus balanoides	FJ845819
Semibalanus balanoides	FJ845818
Semibalanus balanoides	GQ328964
Austrobalanus imperator	EU423232
Tetraclita kuroshioensis	JX186409
Amphibalanus improvisus	FJ845841
Amphibalanus improvisus	FJ845844
Amphibalanus improvisus	FJ845842
Verruca stroemia	JX083863
Catophragmus imbricatus	JX083864
Microeuraphia rhizophorae	FJ845863
Chthamalus panamensis	FJ857976
Chthamalus panamensis	FJ857974
Chthamalus panamensis	FJ857960
Chthamalus panamensis	FJ857966
Chthamalus hedgecocki	FJ857986
Chthamalus hedgecocki	FJ857987
Chthamalus hedgecocki	FJ857990
Chthamalus southwardorum complex sp B	FJ857992
Chthamalus southwardorum complex sp B	FJ857997
Chthamalus southwardorum complex sp B	FJ857999
Chthamalus southwardorum complex sp B	FJ857993
Chthamalus fragilis	AF234807
Chthamalus fragilis	AF234813
Chthamalus anisopoma	AF234816
Chthamalus proteus	FJ858023
Chthamalus proteus	FJ858025
Chthamalus proteus	FJ858039
Chthamalus proteus	FJ858027
Chthamalus mexicanus	AF234804
Chthamalus mexicanus	AF234805
Chthamalus mexicanus	AF234803
Capitulum mitella	JX502999
Chthamalus hedgecocki	FJ857989
Fistulobalanus albicostatus	JX503003
Chthamalus anisopoma	AF234818
Chthamalus anisopoma	AF234817
Chthamalus anisopoma	AF234819
Microeuraphia depressa	HQ224881
Chthamalus stellatus	EU699240
Chthamalus stellatus	EU699243
Microeuraphia withersi	AY430814
Chamaesipho brunnea	AY430811
Jehlius cirratus	GU126095
Jehlius cirratus	GU126107
Jehlius cirratus	GU126087
Jehlius cirratus	GU126116
Octomeris angulosa	AY428049
Notochthamalus scabrosus	FJ845821
Notochthamalus scabrosus	FJ845822
Notochthamalus scabrosus	GU125776
Pollicipes polymerus	GU442485
Pollicipes polymerus	GU442491
Pollicipes polymerus	GU442492
Pollicipes polymerus	GU442494
Pollicipes caboverdensis	HM563665
Pollicipes caboverdensis	HM563666
Pollicipes caboverdensis	HM563667
Pollicipes caboverdensis	HM563668
Pollicipes pollicipes	HM563675
Pollicipes pollicipes	HM563676
Pollicipes pollicipes	HM563678
Pollicipes pollicipes	HM563677
Galkinia adamanteus	JX983109
Octomeris brunnea	AY430812
Galkinia decima	JQ946213
Cantellius septimus	KC138455
Tetraclita pacifica	DQ363694
Tetraclita pacifica	DQ363692
Tetraclita pacifica	DQ363693
Tetraclita pacifica	DQ363695
Tetraclita squamosa	DQ363704
Tetraclita squamosa	DQ363705
Tetraclita squamosa	DQ363703

Tetraclita squamosa	DQ363706
Conopea calceola	HQ290142
Conopea calceola	HQ290143
Conopea sp B DCS2011	HQ290138
Conopea sp B DCS2011	HQ290141
Conopea sp B DCS2011	HQ290139
Conopea sp B DCS2011	HQ290140
Conopea calceola	HQ290134
Conopea sp A DCS2011	HQ290135
Conopea sp A DCS2011	HQ290136
Conopea sp A DCS2011	HQ290137
Conopea cf galeata USA DCS2011	HQ290146
Conopea cf galeata Panama DCS2011	HQ290130
Conopea cf galeata Brazil DCS2011	HQ290133
Conopea cf galeata Galapagos DCS2011	HQ290144
Chthamalus moro	HM135959
Chthamalus moro	HM135960
Tetraclita stalactifera	JN589833
Chelonibia caretta	JN589810
Chelonibia caretta	JN589812
Chelonibia caretta	JN589811
Chthamalus dalli	AY795282
Chthamalus dalli	AY795283
Chthamalus dalli	AY795285
Chthamalus dalli	AY795284
Chthamalus fissus	DQ538424
Chthamalus fissus	DQ538422
Chthamalus fissus	DQ538421
Chthamalus fissus	DQ538423
Chthamalus sp kino	DQ538449
Chthamalus sp kino	DQ538448
Chthamalus sp kino	DQ538447
Chthamalus sp kino	DQ538446
Tetraclita ehsani	JX186296
Tetraclita ehsani	JX186295
Tetraclita ehsani	JX186297
Tetraclita ehsani	JX186298
Tetraclita serrata	JX186199
Tetraclita serrata	JX186201
Tetraclita serrata	JX186200
Tetraclita serrata	JX186202
Tetraclita sp n LMT2012	JX186368
Tetraclita sp n LMT2012	JX186366
Tetraclita sp n LMT2012	JX186369
Tetraclita sp n LMT2012	JX186367
Tetraclita reni	JX186294
Tetraclita reni	JX186365
Tetraclita achituvi	JX186290
Tetraclita achituvi	JX186291
Tetraclita achituvi	JX186289
Tetraclita achituvi	JX186288
Lepas anserifera	KC138475
Lepas pectinata	KC138476
Verruca laevigata	JX083862
Armatobalanus allium	KC138449
Rostratoverruca krugeri	AB195609
Verruca sp CJS2008	EU439973
Brochiverruca sp KT0203a	AB195608
Altiverruca sp NT0207&08	AB195607
Metaverruca recta	JX083861
Octolasmis angulata	KC138498
Octolasmis cor	KC138499
Octolasmis cor	KC138500
Octolasmis orthogonia	EU884173
Octolasmis warwickii	KC138501
Neoverruca brachylepadoformis	AB195606
Neoverruca sp Ok8	AB195598
Neoverruca sp Ok14	AB195604
Neoverruca sp Ok6	AB195596
Neoverruca sp Ok9	AB195599
Neoverruca sp Ok1	AB195591
Neoverruca sp Ok5	AB195595
Neoverruca sp Ok15	AB195605
Neoverruca sp Ok7	AB195597
Neoverruca sp Ok11	AB195601
Neoverruca sp Ok13	AB195603
Neoverruca sp Ok10	AB195600
Neoverruca sp Og5	AB195586
Neoverruca sp Og1	AB195582
Neoverruca sp Og4	AB195585
Neoverruca sp Og3	AB195584
Neoverruca sp Og6	AB195587
Neoverruca sp Og7	AB195588
Neoverruca sp Og9	AB195590
Neoverruca sp Og8	AB195589
Neoverruca sp Ok4	AB195594
Neoverruca sp Ok12	AB195602
Neoverruca sp Ok3	AB195593
Neoverruca sp Og2	AB195583
Neoverruca sp Ok2	AB195592
Chthamalus sp 2 JDZ2005	AY823028
Chthamalus sp 2 JDZ2005	AY823030
Chthamalus sp 2 JDZ2005	AY823029
Chthamalus sp 2 JDZ2005	AY823031
Pollicipes sp JQ2009	GQ472625
Pollicipes sp JQ2009	GQ472627
Pollicipes sp JQ2009	GQ472628
Pollicipes sp JQ2009	GQ472626
Pollicipes elegans	GQ472619
Pollicipes elegans	GQ472614
Pollicipes elegans	GQ472615
Pollicipes elegans	GQ472618
Notochthamalus scabrosus	FJ845824
Calantica spinosa	AY428047
Notochthamalus scabrosus	FJ84582
Calantica spinosa	AY428047

28S

Taxon ID	GenBank Accession		
Altiverruca sp KACb00436	EU082300	*Microeuraphia depressa*	JX083927
Metaverruca recta	EU082297	*Octomeris brunnea*	JX083917
Verruca laevigata	EU082296	*Hexechamaesipho pilsbryi*	JX083918
Verruca stroemia	AY520615	*Pachylasma japonicum*	AB723931
Rostratoverruca sp KACb00435	EU082298	*Stephanolepas muricata*	AB723918
Rostratoverruca krugeri	EU082299	*Tetraclitella divisa*	AY520603
Chelonibia caretta	AB723915	*Tetraclitella purpurascens*	AY520604
Chelonibia manati	AB723917	*Semibalanus balanoides*	AY520592
Chelonibia patula	EU082295	*Semibalanus balanoides*	EU370440
Megabalanus occator	AB723916	*Semibalanus cariosus*	AY520593
Megabalanus tintinnabulum	AY520597	*Microeuraphia rhizophorae*	JX083929
Megabalanus californicus	AY520598	*Lithotrya valentiana*	EU082301
Megabalanus californicus	AY859588	*Lithotrya sp KACb00393*	EU082302
Austromegabalanus psittacus	AY520600	*Balanus crenatus*	AY520590
Megabalanus spinosus	AY520599	*Balanus glandula*	AY520591
Balanus perforatus	AY520595	*Balanus balanus*	AY520594
Menesiniella aquila	AY520596	*Pollicipes pollicipes*	AY520616
Stomatolepas praegustator	AB723919	*Pollicipes pollicipes*	EU370441
Stomatolepas sp RH2012	AB723921	*Lepas sp Lep1*	EU914256
Stomatolepas transversa	AB723920	*Pollicipes polymerus*	AY520617
Tubicinella cheloniae	AB723922	*Capitulum mitella*	AY520618
Platylepas decorata	AB723923	*Calantica spinosa*	EU082303
Cylindrolepas sinica	AB723926	*Calantica sp KACb00087*	EU082304
Xenobalanus globicipitis	AB723927	*Smilium peronii*	EU082305
Coronula diadema	AB723928	*Litoscalpellum discoveryi*	EU489824
Cryptolepas rhachianecti	AB723929	*Trianguloscalpellum regium*	EU082308
Tetraclita squamosa	AY520605	*Litoscalpellum regina*	AY520619
Tetraclita japonica	AY520606	*Scalpellum scalpellum*	EU082307
Platylepas hexastylos	AB723924	*Arcoscalpellum sp CJS2008*	EU489828
Cylindrolepas darwiniana	AB723925	*Arcoscalpellum sp CJS2008*	EU489829
Austrobalanus imperator	AB723930	*Arcoscalpellum africanum*	EU489831
Catomerus polymerus	AY520614	*Arcoscalpellum africanum*	EU489834
Catophragmus imbricatus	JX083911	*Arcoscalpellum africanum*	EU489835
Pseudoctomeris sulcata	JX083912	*Arcoscalpellum africanum*	EU489833
Elminius modestus	AY520601	*Litoscalpellum sp CJS2008*	EU489837
Elminius kingii	AY520602	*Litoscalpellum sp CJS2008*	EU489839
Notochthamalus scabrosus	AY520612	*Scalpellum sp CJS2008*	EU489830
Octomeris angulosa	JX083916	*Ornatoscalpellum stroemii*	EU082306
Chamaesipho brunnea	JX083915	*Arcoscalpellum beuveti*	EU489827
Chamaesipho tasmanica	AY520613	*Leucolepas longa*	EU082311
Chamaesipho sp MPL2012	JX083914	*Vulcanolepas sp KACb00419*	EU082312
Nesochthamalus intertextus	JX083919	*Vulcanolepas osheai*	EU082313
Chamaesipho columna	JX083913	*Neolepas rapanuii*	EU082309
Chthamalus stellatus	AY520607	*Neolepas zevinae*	EU082310
Chthamalus montagui	AY520608	*Ashinkailepas seepiophila*	EU082314
Chthamalus challengeri	AY520609	*Neoverruca sp KACb00361*	EU082315
Jehlius cirratus	AY520611	*Neoverruca sp KACb00389*	EU082316
Microeuraphia withersi	JX083928	*Neoverruca brachylepadoformis*	EU082317
Chthamalus bisinuatus	AY520610	*Ibla quadrivalvis*	AY520621
Tetrachthamalus oblitteratus	JX083925	*Ibla cumingi*	EU082332
Caudoeuraphia caudata	JX083926	*Poecilasma inaequilaterale*	AY520620
Chthamalus malayensis	JX083922	*Poecilasma kaempferi*	EU082329
Chthamalus dentatus	JX083923	*Megalasma striatum*	EU082330
Chthamalus anisopoma	JX083924	*Octolasmis cor*	EU082326
Chthamalus antennatus	JX083920	*Octolasmis sp KACb00064*	EU082327

Octolasmis warwickii	EU082328
Conchoderma auritum	EU082320
Conchoderma virgatum	EU082321
Lepas pectinata	EU082322
Lepas anserifera	EU082323
Lepas australis	EU082324
Lepas testudinata	EU082325
Paralepas dannevigi	EU082318
Oxynaspis celata	EU082331
Heteralepadomorpha sp KACb00398	EU082319

H3

Taxon ID	GenBank Accession
Vulcanolepas sp KACb00419	EU082352
Vulcanolepas osheai	EU082353
Neolepas zevinae	EU082350
Leucolepas longa	EU082351
Neolepas rapanuii	EU082349
Ashinkailepas seepiophila	EU082354
Vulcanolepas sp East Scotia Ridge	JN628252
Neoverruca sp KACb00389	EU082356
Neoverruca sp KACb00361	EU082355
Neoverruca brachylepadoformis	EU082357
Lithotrya sp KACb00393	EU082343
Lithotrya valentiana	EU082342
Tesseropora rosea	JX186507
Tetraclita achituvi	JX186499
Tetraclita rufotincta	JX186501
Tetraclita sp n LMT2012	JX186504
Tetraclita sp n LMT2012	JX186502
Tetraclita sp n LMT2012	JX186505
Tetraclita sp n LMT2012	JX186503
Tetraclita ehsani	JX186500
Tetraclita kuroshioensis	JX186506
Tetraclita serrata	JX186497
Tetraclita serrata	JX186496
Tetraclita serrata	JX186498
Trianguloscalpellum regium	EU082348
Chelonibia patula	EU082337
Metaverruca recta	EU082339
Verruca laevigata	EU082338
Conopea calceola	HQ290156
Conopea calceola	HQ290155
Conopea calceola	KC349910
Conopea fidelis	KC349914
Conopea sp B DCS2011	HQ290150
Conopea fidelis	KC349912
Conopea sp B DCS2011	HQ290158
Conopea sp B DCS2011	HQ290157
Conopea fidelis	KC349905
Conopea sp B DCS2011	HQ290151
Conopea calceola	HQ290152
Conopea calceola	HQ290149
Conopea sp A DCS2011	HQ290160
Conopea saotomensis	KC349911
Conopea saotomensis	KC349906
Conopea saotomensis	KC349913
Conopea saotomensis	KC349909
Conopea saotomensis	KC349904
Conopea sp A DCS2011	HQ290159
Conopea saotomensis	KC349907
Conopea saotomensis	KC349908
Rostratoverruca krugeri	EU082341
Rostratoverruca sp KACb00435	EU082340
Megalasma striatum	EU082370
Octolasmis sp KACb00064	EU082367
Lepas anserifera	EU082363
Lepas australis	EU082364
Lepas testudinata	EU082365
Lepas pectinata	EU082362
Octolasmis cor	EU082366
Oxynaspis celata	EU082371
Heteralepadomorpha sp KACb00398	EU082359
Octolasmis warwickii	EU082368
Scalpellum scalpellum	EU082347
Calantica sp KACb00087	EU082345
Smilium peronii	EU082346
Calantica spinosa	EU082344
Pyrgopsella sp SMB2011	JN800715
Savignium crenatum	JN800716
Conchoderma auritum	EU082360
Conchoderma virgatum	EU082361
Conopea cf galeata Galapagos DCS2011	HQ290162
Conopea cf galeata Galapagos DCS2011	HQ290161
Conopea cf galeata USA DCS2011	HQ290147
Conopea galeata	JQ966286
Conopea galeata	JQ966283
Conopea galeata	JQ966284
Conopea galeata	JQ966285
Paralepas dannevigi	EU082358
Pollicipes elegans	HM142348
Pollicipes sp RJVS2010	HM142349
Poecilasma kaempferi	EU082369
Ibla cumingi	EU082372

Table S3. Predictions of number of RAD-tags in thoraciacan barnacles using SbfI. Data for *Daphnia pulex* obtained from the U.S. National Center for Biotechnology Information (NCBI) WGS database. Observed frequency of recognition sequences and calculated probability based on a trinucleotide genome composition model were generated following the methodology described by Herrera *et al.*(Herrera *et al.* 2014). Data for known barnacle genome sizes obtained from the Animal Genome Size Database (http://ww.genomesize.com). C-value is the amount of DNA in picograms in the nucleous, where the genome size in Mbp = 978 x C-value

Species	**Common name**	**C-value**	**Genome size (Mbp)**	**Observed frequency of SbfI recognition sites per bp**	**Probability of SbfI recognition site per bp**
Genebank WGS					
Daphnia pulex	Water flea		158.61	7.48E-06	5.90E-06

Species	**Common name**	**C-value**	**Genome size (Mbp)**	**Predicted number of SbfI recognition sites based on *D. pulex* observed frequency**	**Predicted number of SbfI RAD-tags based on *D. pulex* observed frequency**	**Predicted number of SbfI recognition sites based on *D. pulex* trinucleotide genome composition probability**	**Predicted number of SbfI RAD-tags based on *D. pulex* trinucleotide genome composition probability**
Animal Genome Size Database							
Balanus amphitrite	Striped barnacle	0.74	723.72	5,411.68	10,823.35	4,267.22	8,534.43
Balanus amphitrite	Striped barnacle	1.4	1,369.20	10,238.31	20,476.61	8,073.11	16,146.22
Balanus cariosus	Thatched barnacle	1.4	1,369.20	10,238.31	20,476.61	8,073.11	16,146.22
Balanus eburneus	Ivory barnacle	1.26	1,232.28	9,214.48	18,428.95	7,265.80	14,531.60
Chthalamus sp.	Acorn barnacle	1.23	1,202.94	8,995.08	17,990.17	7,092.80	14,185.61
Unknown sp.	Goose barnacle	1.46	1,427.88	10,677.09	21,354.18	8,419.10	16,838.20
Mitella polymerus	Pacific goose barnacle	0.9	880.20	6,581.77	13,163.54	5,189.86	10,379.71
Tetraclita rubescens	Volcano barnacle	2.6	2,542.80	19,014.00	38,027.99	14,992.92	29,985.84

Table S4. Nucleotide substitution models for each Sanger-based genetic marker, as selected by the BIC criterion in jModeltest.

Dataset	coxI_codon12	H3_codon12	H3_codon3	28S
Thoracica	GTR+I+G	HKY+I+G		GTR+I+G
Clade A	HKY+I	HKY+I		HKY+G

Table S5. Results from Xia saturation test for each Sanger-based genetic marker.

Dataset	coxI_codon12	H3_codon12	H3_codon3	28S
Thoracica	Little saturation	Substantial saturation	Little saturation	Little saturation
C1	Little saturation	Substantial saturation	Little saturation	Little saturation

Table S6. RAD sequencing results and filtering statistics.

				STACKS filtering			pyRAD filtering		
ID Species	**ID sample**	**DNA integrity, HMW band intensity, and shear level (agarose gel)***	**Total sequenced reads**	**Reads discarded due to absence of RAD-tag**	**Reads discarded due to low quality**	**Retained reads**	**Reads that passed quality filtering**	**Total number of retained reads**	**Percentage of retained reads after filtering steps**
Vulcanolepas_scotiaensis_ESR_E2	16640	H; Strong HMW; NS	1,079,806.00	45,540.00	37,560.00	996,706.00	955,965.00	955,965.00	88.53
Vulcanolepas_osheai_Kermadec_Healy	72638_22	H; Strong HMW; NS	1,647,804.00	35,100.00	64,050.00	1,548,654.00	1,476,024.00	1,476,024.00	89.58
Vulcanolepas_osheai_Kermadec_Tangaroa	82121_15	H; Strong HMW; LS	2,133,810.00	47,251.00	95,743.00	1,990,816.00	1,875,914.00	1,875,914.00	87.91
Ashinkailepas_kermadecensis_Lau_Basin_Niua_N	AsNiN2	H; Strong HMW; LS	1,091,017.00	43,293.00	41,739.00	1,005,985.00	954,951.00	954,951.00	87.53
Ashinkailepas_seepiophila_Okinawa_Trough	AsOk3	M; Moderate HMW, SS	580,247.00	107,879.00	14,673.00	457,695.00	438,974.00	438,974.00	75.65
Neolepas_zevinae_EPR_Tica2	bar22	M; Moderate HMW, SS	643,213.00	53,267.00	20,219.00	569,727.00	549,871.00	549,871.00	85.49
Neobrachylepas_relica_Lau_Basin_NELSC	bar06	H; Strong HMW; LS	1,004,066.00	69,318.00	35,161.00	899,587.00	856,092.00	856,092.00	85.26
Neolepas_sp1_SWIR_Dragon_vent	JC6731B11	H; Strong HMW; LS	984,409.00	65,379.00	27,756.00	891,274.00	862,579.00	862,579.00	87.62
Neolepas_sp1_CIR	NeIn2	L; No HMW, SS	232,285.00	142,801.00	6,243.00	83,241.00	76,759.00	76,759.00	33.05
Neoverruca_sp1_Ogasawara_Arc	NeOg1	L; No HMW, SS	104,531.00	16,470.00	5,000.00	83,061.00	77,248.00	77,248.00	73.90
Neolepas_zevinae_SEPR	SEPR3	L; No HMW, SS	236,557.00	70,581.00	7,952.00	158,024.00	149,850.00	149,850.00	63.35
Vulcanolepas_sp1_Lau_Basin_Mata_Ua	VuMaU1	H; Strong HMW; LS	923,322.00	55,368.00	41,035.00	826,919.00	783,374.00	783,374.00	84.84
Leucolepas_longa_Mariana_Arc_TOTO_vent	VuTO2	M; Moderate HMW, SS	304,975.00	87,453.00	7,744.00	209,778.00	202,382.00	202,382.00	66.36

*Abbreviations: High integrity (H), Medium integrity (M), Low integrity (L), High Molecular Weight (HMW), No Shear (NS), Low Shear (LS), Strong Shear (SS)

Table S7. RAD clustering statistics.

pyRAD clustering (c80)

ID sample	Total number of clusters	Mean depth of clusters	Standard deviation of cluster depth	Number of clusters with depth greater than 5	Mean depth of clusters with depth greater than 5	Standard deviation of clusters with depth greater than 5
16640	21,256	43.61	127.54	18,198	50.62	136.60
72638_22	25,837	54.66	219.37	22,289	63.05	235.10
82121_15	26,882	66.65	279.28	23,015	77.53	300.47
AsNiN2	22,793	38.14	189.94	18,847	45.70	208.09
AsOk3	17,057	24.22	131.27	12,610	32.08	151.89
bar22	18,745	28.32	88.47	14,987	34.93	97.83
bar06	20,640	39.36	160.01	17,346	46.47	173.64
JC6731B11	20,340	40.74	148.45	17,359	47.39	159.75
NeIn2	6,133	12.25	25.82	3,706	19.07	31.39
NeOg1	6,055	12.36	18.57	3,797	18.61	21.09
SEPR3	11,611	12.73	24.74	7,380	18.96	29.25
VuMaU1	20,884	35.76	95.06	17,346	42.65	102.96
VuTO2	10,607	18.50	59.61	7,508	25.33	69.71

pyRAD clustering (c85)

ID sample	Total number of clusters	Mean depth of clusters	Standard deviation of cluster depth	Number of clusters with depth greater than 5	Mean depth of clusters with depth greater than 5	Standard deviation of clusters with depth greater than 5
16640	19,884	44.17	101.89	17,283	50.52	107.86
72638_22	24,008	57.15	228.03	21,070	64.84	242.41
82121_15	24,580	69.94	282.61	21,537	79.55	300.68
AsNiN2	21,056	39.13	191.94	17,750	46.05	208.32
AsOk3	15,728	24.50	105.61	11,915	31.73	120.45
bar22	17,698	28.61	83.40	14,291	34.96	91.68
bar06	19,441	40.50	155.99	16,580	47.15	168.02
JC6731B11	19,045	41.52	139.60	16,444	47.77	149.28
NeIn2	5,848	12.63	27.34	3,600	19.35	33.10
NeOg1	5,831	12.76	19.88	3,712	18.99	22.66
SEPR3	11,244	12.81	20.59	7,195	18.96	23.60
VuMaU1	19,413	36.77	94.33	16,375	43.23	101.40
VuTO2	10,045	18.81	59.55	7,194	25.49	69.23

pyRAD clustering (c90)

ID sample	Total number of clusters	Mean depth of clusters	Standard deviation of cluster depth	Number of clusters with depth greater than 5	Mean depth of clusters with depth greater than 5	Standard deviation of clusters with depth greater than 5
16640	19,181	44.74	103.33	16,828	50.72	108.99
72638_22	23,039	57.73	211.75	20,417	64.89	223.93
82121_15	23,544	71.10	287.50	20,872	79.96	304.21
AsNiN2	20,266	39.13	174.49	17,184	45.78	188.72
AsOk3	15,095	24.79	106.79	11,585	31.71	121.05
bar22	17,166	28.80	84.49	13,943	35.00	92.65
bar06	18,912	41.06	158.41	16,206	47.59	170.25
JC6731B11	18,387	41.72	137.37	16,015	47.61	146.27
NeIn2	5,679	12.66	25.03	3,542	19.18	29.85
NeOg1	5,692	12.95	20.34	3,661	19.11	23.16
SEPR3	11,021	12.86	19.55	7,089	18.95	22.13
VuMaU1	18,692	37.02	94.83	15,904	43.16	101.57
VuTO2	9,761	19.05	60.22	7,043	25.65	69.78

Table S8. RAD-seq matrices statistics

RAD-seq matrix	Number of loci	Number of base pairs	Percentage of missing data	Number of variable sites	Percentage of variable sites	Number of parsimony informative sites	Percentage of parsimony informative sites
c80d5m4p3	15,331	1,303,485	51.15	121,475	9.32	37,722	2.89
c80d5m6p3	9,824	835,710	43.18	85,987	10.29	29,077	3.48
c80d5m8p3	4,039	343,583	33.12	39,877	11.61	14,084	4.10
c80d5m10p3	675	57,406	20.91	7,564	13.18	2,423	4.22
c85d5m4p3	15,499	1,314,995	51.62	109,124	8.30	35,303	2.68
c85d5m6p3	9,766	828,960	43.54	76,353	9.21	26,955	3.25
c85d5m8p3	3,884	329,620	33.37	33,894	10.28	12,689	3.85
c85d5m10p3	618	52,373	21.17	5,984	11.43	2,041	3.90
c90d5m4p3	15,595	1,318,652	52.78	89,781	6.81	29,801	2.26
c90d5m6p3	9,310	787,167	44.29	59,659	7.58	21,828	2.77
c90d5m8p3	3,396	286,979	33.74	23,903	8.33	9,559	3.33
c90d5m10p3	481	40,605	21.38	3,592	8.85	1,357	3.34

CHAPTER 4

No evidence of seamount-driven isolation in deep-sea hydrothermal vent barnacle populations

ABSTRACT

Patterns of spatial genetic population structuring provide insight into the factors that limit dispersal and connectivity of species. Deep-sea hydrothermal vents are the focus of increasing interest for the mining of mineral resources found in polymetallic sulphide deposits. Understanding the genetic diversity and population connectivity of vent species is critical for assessing the potential effects of mining on these ecosystems. Population genetic studies of vent species have mainly focused on mid-ocean ridge systems and back-arc spreading centers. However, vents also occur in active seamounts worldwide. Seamounts are hypothesized to behave as isolated island-like systems, where population connectivity is limited and endemicity is promoted (seamount endemicity hypothesis). In this study, we aim to test this seamount endemicity hypothesis using novel genome-wide restriction-site associated DNA (RAD) sequence data from three hydrothermal vent barnacle species. Comparisons of the genetic diversity and population structuring patterns of barnacle populations from seamounts and spreading ridges revealed patterns of population genetic structuring that do not conform to the predictions from the seamount endemicity hypothesis. The patterns of genetic variation among individuals collected from seamounts and spreading ridges, separated horizontally by hundreds of kilometers and vertically by hundreds of meters, did not reject the null hypothesis of panmixia within each species. These inferences are largely insensitive the to *de novo* assembly parameters used to identify loci from sequence reads. We suggest that the seamount endemicity hypothesis warrants further testing using high-resolution genetic markers in other vent organisms with differing life history strategies (e.g., brooders) that may limit their dispersal potential, as well as in non-vent organisms, which are not exposed to evolutionary pressures imposed by the dynamic nature of hydrothermal vent systems.

INTRODUCTION

Patterns of spatial genetic population structuring provide insight into the factors that limit dispersal and connectivity throughout a species' range. Empirical evidence has revealed the fundamental importance of intrinsic and extrinsic factors, such as habitat discontinuity (Reitzel *et al.* 2013; D'Aloia *et al.* 2014), depth (Prada & Hellberg 2013), currents (Kelly & Palumbi 2010; White *et al.* 2010), distance (Alberto *et al.* 2011), larval developmental mode (Kelly & Palumbi 2010), and symbiotic associations (Beinart *et al.* 2012), acting as barriers to gene flow in marine species and ultimately driving biodiversity patterns in the ocean. Understanding how populations are spatially and temporally interconnected is critical due to the need for marine biodiversity and ecosystems protection against increasing threats created by human activities (Christensen *et al.* 1996).

Deep-sea hydrothermal vents are the focus of increasing interest for the mining of mineral resources found in polymetallic sulphide deposits (Boschen *et al.* 2013), which form by precipitation during mixing of metal-rich vent fluids with bottom seawater. Deep-sea hydrothermal vents (hereafter simply referred to as vents) host some of the most spectacular and unique ecosystems on earth that thrive on *in situ* primary productivity derived from chemosynthesis. Vents present a sharp contrast to other ecosystems in the deep sea, due to their marked patchiness, extremely steep chemical and thermal gradients over centimeter scales, and relatively high frequency of disturbances given their occurrence on highly dynamic geological settings (Van Dover 2000). Consequently, vent environments present extreme selective pressures, evolutionary speaking, and are characterized by their low biodiversity and high endemicity. These characteristics make vent ecosystems potentially susceptible to disturbances caused by deep-sea mining. Although vent organisms have adapted to the natural dynamics and ephemerality of hydrothermal vents, the potential disturbances from mining are likely to have multiplicative harmful effects at unprecedented scales (Van Dover 2010). Understanding the genetic diversity and population connectivity of vent species is critical to assessments of the potential effects of mining on these ecosystems.

Population genetic studies of vent species from spreading ridge systems commonly reveal patterns of genetic diversity consistent with high gene flow along ridge axes extending for hundreds to thousands of kilometers, yet factors such as depth and ocean currents have been identified as barriers to dispersal at regional scales (see review by Vrijenhoek (2010)). The geomorphology of spreading ridges is known to modify local current regimes and generate significant current flows along ridge axes (Thurnherr *et al.* 2011; Lavelle *et al.* 2012), which are hypothesized to facilitate dispersal among spatially separated vent fields over long distances. Deep-sea hydrothermal vents also occur on active volcanic seamouts located

on hotspots or island arcs. However, the population structuring and genetic connectivity dynamics in these systems are much less understood.

Seamounts are hypothesized to behave as isolated island-like systems, where population connectivity, and therefore gene flow, are limited (commonly referred to as the 'seamount endemicity hypothesis', de Forges *et al.* 2000). Such isolation might arise due to the patchiness of the seamounts habitats (particularly of active seamounts as most seamounts are inactive), combined with larval behaviors, physical oceanographic phenomena, such as Taylor columns, that could cause larvae retention around seamount summits (Wilson & Kaufmann 1986; Mullineaux & Mills 1997; de Forges *et al.* 2000; Metaxas 2011), or a combination of these factors. The seamount endemicity hypothesis predicts that there will be relatively high levels of endemicity of genetic variants and significant structure among populations from different seamounts. Consistent with this hypothesis Smith *et al.* (2004) found heterogeneity in allozyme frequencies and significant genetic differentiation between populations of *Bathymodiolus* vent mussels from two active seamounts of the Kermadec arc. Contrastingly, Tunnicliffe *et al.* (2010) and Watanabe *et al.* (2005) found significant differentiation in mitochondrial sequences between volcanic arc basins, but no differentiation among seamount populations of vent flatfish and barnacle species within each arc basin. Other studies in non-vent seamount fauna have found similar patterns of absence of population structuring among seamount populations of deep-sea corals (Thoma *et al.* 2009), clams (Clague *et al.* 2012), and ophiuroids (Cho & Shank 2010), when examining mitochondrial markers. However, the mitochondrion has a intermediate mutation rate compared to other autosomal markers such as microsatellites and single nucleotide polymorphisms (SNPs), and mounting evidence indicates that significant amounts of genetic variability in populations and species can go undetected based solely on mitochondrial data (Emerson *et al.* 2010; Herrera & Shank Chapter 5). For example, a study on a species of Patagonian toothfish revealed significant genetic structure between eastern and western Southern Ocean populations using data from microsatellites, when a lack of genetic structure had been previously concluded using mitochondrial data only (Rogers *et al.* 2006). Therefore, interpretations of non-endemicity in seamounts based on mitochondrial DNA data have been challenged (Baco & Cairns 2012).

In this study we aim to test the seamount endemicity hypothesis in deep-sea hydrothermal-vent fauna using novel data from thousands of genome-wide SNPs obtained from restriction-site associated DNA (RAD) markers (Baird *et al.* 2008). Our approach is to compare the genetic diversity and population structuring patterns of seamount and spreading ridge populations in three hydrothermal vent barnacle species: 1) *Vulcanoleapas scotiaensis* from East Scotia ridge; 2) *Vulcanolepas osheai* from the Kermadec arc; and 3) *Eochionelasmus ohtai* from the Lau Basin. Under the seamount endemicity hypothesis, we

expect to find higher population genetic structuring between seamount populations than between populations from spreading ridges.

METHODS

Barnacle pecimens were collected from deep-sea hydrothermal vents at 2-3 localities each within the East Scotia ridge (E2 and E9 vent sites), the Kermadec arc (Healy and Tangaroa seamounts), and the Lau Basin (Tahi Moana vent site, and the Mata Ua and Founalei South seamounts). Morphological identifications were performed on using stereo-microscopy and species descriptions as references (Table 1, Fig. 1). Species were identified as *Vulcanoleapas scotiaensis, Vulcanolepas osheai*, and *Eochionelasmus ohtai* respectively. Species identifications of representative specimens were confirmed through multilocus phylogenetic analyses by Herrera *et al.* (Chapter 3).

Table 1. Collection information for the specimens used in this study.

Species	Region	Population locality	Collection event	Depth (m)	Lat.	Lon.	Collection ID
Vulcanolepas scotiaensis	East Scotia ridge	E2 vent field	JC042	2700	-56.06	-30.33	WHOI-NEOE2
Vulcanolepas scotiaensis	East Scotia ridge	E9 vent field	JC042	2400	-60.05	-29.93	WHOI-NEOE9
Vulcanolepas osheai	Kermadec arc	Healy volcano	TAN1104/073	1255	-35.01	178.98	NIWA-72638
Vulcanolepas osheai	Kermadec arc	Tangaroa volcano	TAN1206/017	682	-36.32	178.03	NIWA-82121
Eochionelasmus ohtai	Lau Basin	Tahi Moana vent field	TN236/J2-444	2232	-20.68	-176.18	WHOI-13288
Eochionelasmus ohtai	Lau Basin	Founalei South volcano	RR1211/Q4K-323	956	-17.54	-174.58	WHOI-20665
Eochionelasmus ohtai	Lau Basin	Mata Ua volcano	RR1211/Q4K-328	2391	-15.02	-173.79	WHOI-20791

Molecular laboratory methods

Total genomic DNA was purified from specimens by: (1) digesting the tissue in 2% CTAB buffer (Teknova) with proteinase K and RNAse A/T1 (Fermentas) for 1 hour, (2) separating nucleic acids with chloroform: isoamyl alcohol (24:1) (Fermentas) and phenol: chloroform: isoamyl alcohol (25:24:1, Tris buffered at pH 8.0) (Fermentas), (3) precipitating nucleic acids with 100% ethanol (1:1 volume ratio), and (4) washing the precipitate twice with 70% ethanol.

Concentration-normalized genomic DNA was submitted to Floragenex Inc. (Eugene, OR) for library preparation and RAD sequencing. In short, the RAD sequencing method consists of: 1) the digestion of genomic DNA for each individual with a restriction enzyme; 2) ligation of the resulting fragments to sequencing adapters with unique barcodes for each individual; 3) size-selection and enrichment of the

fragments successfully ligated to the adapters; and 4) sequencing via a high-throughput platform. Individual libraries for *Vulcanolepas osehai* and *Eochionelasmus ohtai* were produced from DNA digested with a high-fidelity SbfI restriction enzyme, which is predicted to cut between 5,000 and 15,000 times in the genome of a thoracican barnacle (Table S1) (Herrera *et al.* Chapter 3). Libraries for *Vulcanolepas scotiaensis* were produced from DNA digested with a high-fidelity SgrAI restriction enzyme, which is predicted to cut between 20,000 and 60,000 times in the genome of a thoracican barnacle. The cut-frequency ranges for both SbfI and SgrAI were predicted by the observed recognition sequence frequencies and their estimated probability using trinucleotide composition models in the genome of the crustacean *Daphnia pulex* using the software predRAD (Herrera *et al.* Chapter 2). Ranges of genome size for barnacles were obtained from the Animal Genome Size Database (http://ww.genomesize.com). Barcode tags were 10-base pairs long. Libraries were sequenced by multiplex on an Illumina Hi-Seq 2000 sequencer.

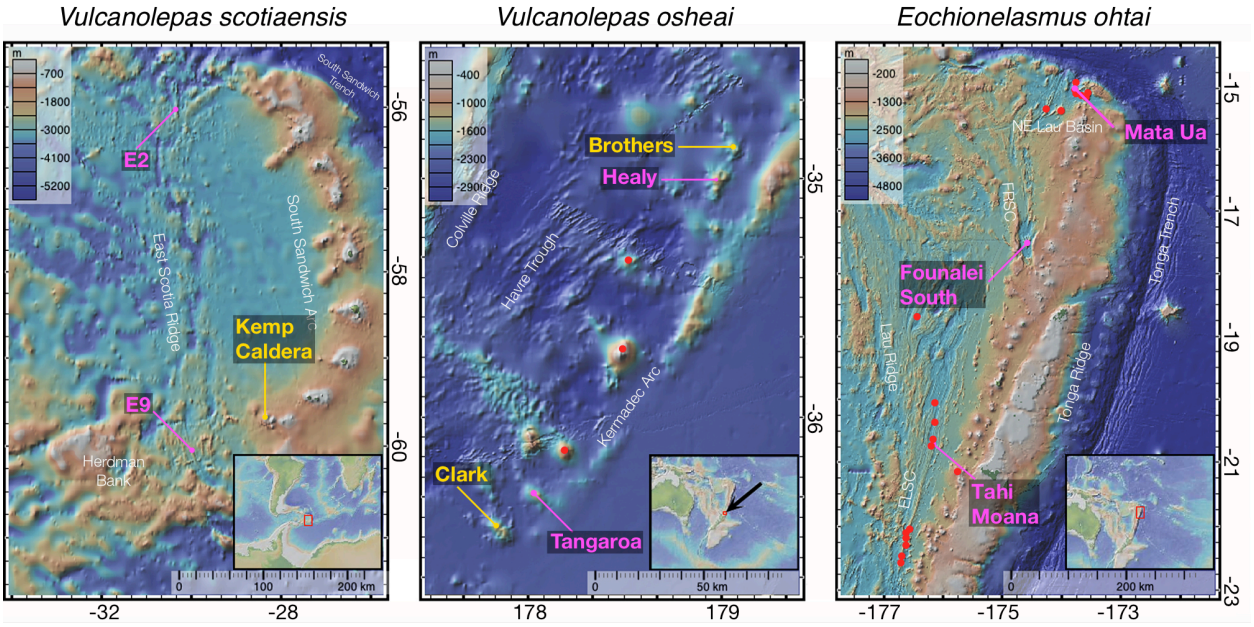


Figure 1. Geographic location of vent barnacle populations for each species. Bathymetry maps show the specific areas where samples have been obtained. Purple labels indicate locations of collections of populations included in this study, but marked to provide context. Yellow labels indicate locations of known populations not included in this study for context. Red dots indicate locations of other active hydrothermal vent sites. Color scale indicates corresponding depths in meters. Bottom right bars indicate distance scale in kilometers. Small maps inset within the bottom right of each subplot indicate region location in a global perspective. Base bathymetry maps were generated using the program GeoMapApp (http://www.geomapapp.org) with data from the Global Multi-Resolution Topography (GMRT) Synthesis (Ryan *et al.* 2009).

Data filtering

Sequence reads were de-multiplexed and quality filtered with the *process_radtags* program from the package Stacks v1.19 (Catchen *et al.* 2013b). Barcodes and Illumina adapters were excluded from each read, and length was truncated to 70bp (-*t* 70). Reads with ambiguous bases were discarded (-*c*). Reads with an average quality score below 10 (-*s* 10) within a sliding window of 15% of the read length (-w *0.15*) were discarded (-*q*). The rescue barcodes and RAD-tags algorithm was enabled (-*r*).

Tests of parameters for de novo loci assembly

To explore the effects of parameter choice on *de novo* assemblies of RAD loci using the *denovo_map* pipeline in Stacks, we tested an array of parameters following the guidelines and modified R scripts by Mastretta-Yanes *et al.* (2014). For each species dataset we independently performed multiple *de novo* loci assemblies modifying the following individual parameter values, while keeping default values for the rest: the minimum number of identical raw reads required to create a stack of identical unique sequences for each individual (-*m* 2, 4, 6, 8, 10 for *V. scotiaensis*; and -*m* 2-8, 10, and 12 for *V. osheai and E. ohtai*), the maximum number of stacks at a single locus for each individual (--*max_locus_stacks* 2-5 for *V. scotiaensis*; and --*max_locus_stacks* 2-6 for *V. osheai and E. ohtai*), the number of mismatches allowed between loci for each individual (-*M* 2, 3, 4 for *V. scotiaensis*; and -*M* 2-6, 8 for *V. osheai and E. ohtai*), and the number of mismatches allowed between loci when building the catalog of all loci in a species (-*n* 0-4 for *V. scotiaensis*; and -*n* 0-5 for *V. osheai and E. ohtai*). High-confidence SNP calls in STACKS are performed using a maximum-likelihood framework that accounts for sequencing error and variable depth of coverage among loci (Hohenlohe *et al.* 2010; Catchen *et al.* 2013b). The results from each assembly were loaded onto a MySQL database and indexed through the *load_radtags* and *index_radtags* tools.

We calculated the number of reads and loci coverage per individual from each assembly, and filtered the data to exclude individuals having less than 50% of the mean number of loci per individual and keep only loci present in at least 80% of the individuals. We executed the *populations* program of Stacks after each *de novo* assembly using only individuals and loci that passed the aforementioned filter (whitelist -*W*). To explore the influence of different assembly parameters on population differentiation estimates, we calculated population F_{ST} values from SNPs utilizing a *p_value* filter (-*f*) to keep only significant estimates (α=0.05). To examine the effect of different assembly parameters on the inferred genetic variability within and among populations, we calculated Euclidean distances among individuals from exported SNP data in *plink* format, and performed Neighbor Joining similarity and principal component analyses in R. Optimal *de novo* assembly parameter settings were chosen conservatively, aiming to

maximize the number of SNPs recovered while minimizing the within-population genetic distances, following guidelines by Mastretta-Yanes *et al.* (2014).

Demographic inferences

We performed *de novo* assemblies of RAD loci using optimal parameter settings for each species in Stacks, as explained above. To estimate population genetic descriptive statistics per SNP (genetic diversity π, proportion of polymorphic loci, observed heterozygosity, minor allele frequency, number of private alleles, inbreeding index $F_{IS,}$ and population differentiation index F_{ST}) we executed the *populations* program of Stacks, using only individuals and loci that passed aforementioned filters. We only analyzed loci that were present in all populations of each species (-*p*) and in at least 75% of individuals per population (-*r*). As before, we calculated population F_{ST} values from SNPs utilizing a *p_value* filter (-*f*) correction (α=0.05). We exported SNP data in *plink* format, keeping only one SNP per loci to avoid violating the assumption of independence among markers. To detect possible population structuring we constructed Neighbor Joining similarity dendrograms and performed principal component analyses with data from individuals for each species, as described above. To explore possible non-equilibrium signals in populations we examined frequency distribution plots of minor allele frequencies and inbreeding indices (F_{IS}) in R.

RESULTS

Sequencing results

We generated RAD-seq data for 117 individuals from three species of vent barnacles: 28 individuals from two populations of *Vulcanolepas scotiaensis,* E2 and E9, from the East Scotia rise; 36 individuals from two populations of *Vulcanolepas osheai*, Healy and Tangaroa, from the Kermadec arc; and 53 individuals from three populations of *Eochionelasmus ohtai,* Tahi Moana, Founalei South, and Mata Ua, from the Lau Basin (Table 2). We obtained approximately 3.1±1.5 (mean ± standard deviation) million reads per individual for *V. scotiaensis*, with individual values ranging from 1.1 to 7.2 million reads. For *V. osheai* we obtained approximately 1.4±0.4 million reads per individual, with individual values ranging from 0.8 to 2.6 million reads. Lastly, for *E. ohtai* we obtained approximately 2.3±0.9 million reads per individual, with individual values ranging from 0.8 to 4.8 million reads. Overall, more than 88% of sequence reads were retained after quality filtering.

Table 2. RAD sequencing results, filtering and *de novo* assembly statistics. N indicates the number of specimens per population. The numbers of raw and retained reads are shown as population mean ± standard deviation.

Species	Population	Sequence IDs	N	Raw reads	Retained reads	% reads retained
Vulcanolepas scotiaensis	E2	NEO_*_E2	14	3,531,674±1,490,665	3,058,341±1,086,450	88±6
Vulcanolepas scotiaensis	E9	NEO_*_E9	14	2,744,428±1,441,426	2,369,012±1,087,019	88±6
Vulcanolepas osheai	Healy	72638_*	18	1,254,447±245,024	1,134,690±249,516	90±6
Vulcanolepas osheai	Tangaroa	82121_*	18	1,500,636±424,632	1,370,937±418,941	91±5
Eochionelasmus ohtai	Tahi Moana	13288_*	17	2,170,591±906,494	1,978,867±876,981	90±5
Eochionelasmus ohtai	Founalei South	EoFo_*	18	2,438,023±728,551	2,322,302±742,900	95±3
Eochionelasmus ohtai	Mata Ua	EoMaU_*	18	2,167,746±1,180,479	2,147,705±1,161,193	91±5

* represents individual IDs

De novo loci assembly parameter examination

De novo loci assembly parameter variation tests produced between 18-55 thousand loci for *V. scotiaensis* containing on average 1.9-2.0 SNPs per locus (Fig. 2). These tests produced 4.1-5.5 and 10.5-13.0 thousand loci, containing on average 2.38-2.43 and 2.17-2.37 SNPs per locus, for *V. scotiaensis* and *E. ohtai,* respectively. Overall, the parameter controlling the minimum number of identical raw reads required to create a stack of identical unique sequences for each individual (*-m*) had the largest influence on the number of markers obtained, particularly in the dataset from *V. scotiaensis* with a roughly linear decrease rate of 5 thousand loci per parameter unit increase (compared to a decrease rate of nearly 200 and 130 loci per parameter unit increase in *V. osheai* and *E. ohtai*, respectively). The parameter *-M*, which limits the number of mismatches allowed between loci for each individual, also produced a general decrease in the number of loci and SNPs as parameter value increased. The parameter *–n*, which limits the number of mismatches allowed between loci when building the catalog of all loci in a species, produced a similar decreasing effect to parameters *-m* and *-M,* after the number of loci and SNPs peaked at *-n* 1. The parameter *--max_locus_stacks*, which limits the maximum number of stacks at a single locus for each individual, did not have a substantial effect on the number of loci or SNPs.

All individuals passed the imposed requirement of having more than 50% of the mean number of loci per individual per species. Mean F_{ST} population differentiation values were consistently low, ranging from 0.0008 to 0.0020 (Fig. 3). Parameter *-n* had the greatest influence in the mean F_{ST} values, which tended to increase and plateau with the parameter values (except in *V. osheai,* where mean F_{ST} values oscillated). Increasing values of parameters *-M* and *-m* had opposite effects to parameter *-n*, causing a general drop in mean F_{ST} values. Parameter *--max_locus_stacks* had negligible effects on the estimates of mean F_{ST} values.

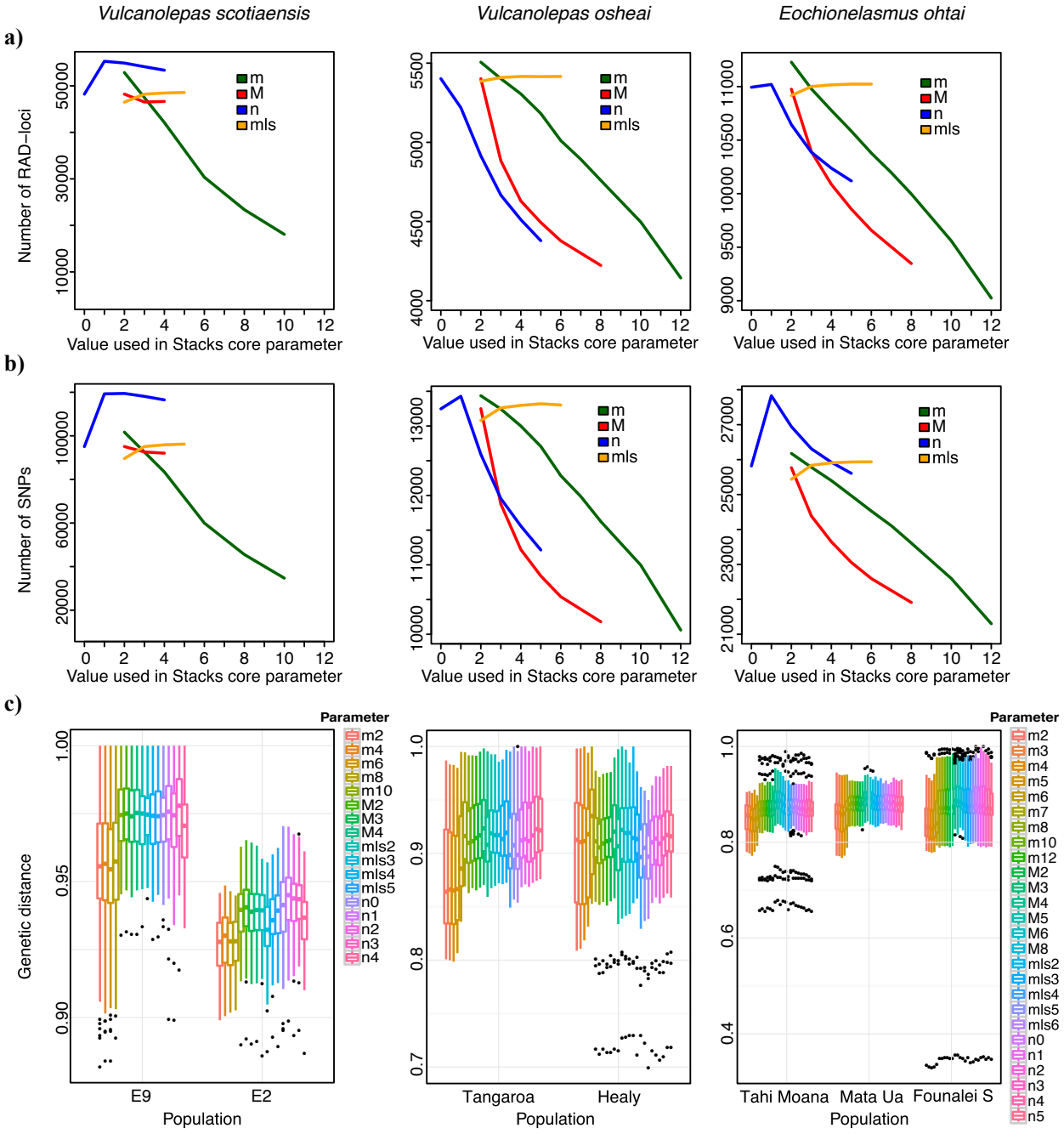


Figure 2. *De novo* loci assembly parameter influence on number of obtained loci, SNPs, and genetic distances for each vent barnacle dataset. Species are indicated at the top of each column. **a)** Plots of the value of each of the examined STACKS core parameters against the number of RAD-loci generated. **b)** Plots of the value of each of the examined STACKS core parameters against the number of SNPs generated. Line colors and legend indicate evaluated core parameters: the minimum number of identical raw reads required to create a stack of identical unique sequences for each individual (-*m* in green), the number of mismatches allowed between loci for each individual (-*M* in red); the number of mismatches allowed between loci when building the catalog of all loci in a species (-*n* in

blue); and the maximum number of stacks at a single locus for each individual (*--max_locus_stacks* abbreviated as mls in yellow). **c)** Box plots of Euclidean genetic distances among individuals, per population, calculated using SNP data from each *de novo* loci assembly parameter examination. Colors and legend indicate evaluated core parameters.

Overall, Euclidean genetic distances among individuals within each population, calculated from SNPs in loci that were present in at least 80% of the individuals for each one of the different *de novo* loci assemblies, were markedly elevated (above 0.8) (Fig. 4). No significant differences in distance distributions were observed among populations or among loci assemblies using different parameters. Neighbor-Joining similarity dendrograms and principal component analyses performed with SNP data from each of the *de novo* loci assemblies were remarkably similar within species, and none of them revealed patterns indicative of population genetic structuring (Fig. 4).

Table 3. Summary statistics of *de novo* RAD stacks assembly. Stacks are defined as clusters of identical reads. N indicates the number of individuals per population. Values per population are shown as mean ± standard deviation.

Species	Population	N	Number of Stacks	Mean coverage depth	S.D. of coverage depth	Maximum coverage depth
Vulcanolepas scotiaensis	E2	14	99,758±10,413	25±5	152±49	28,284±11,158
Vulcanolepas scotiaensis	E9	14	85,349±20,098	23±7	130±56	23,204±13,867
Vulcanolepas osheai	Healy	18	28,533±3,869	38±10	139±34	12,327±3,133
Vulcanolepas osheai	Tangaroa	18	28,993±2,620	44±14	177±62	15,851±5,834
Eochionelasmus ohtai	Tahi Moana	17	28,074±2,449	65±24	324±151	37,552±20,832
Eochionelasmus ohtai	Founalei South	18	27,764±2,524	76±21	386±125	42,658±17,162
Eochionelasmus ohtai	Mata Ua	18	26,504±3,069	71±33	371±198	42,108±25,723

De novo loci assembly with selected parameters

We selected *de novo* assembly parameters conservatively, aiming to prevent pronounced losses of loci and SNPs while ensuring that there were enough reads to make high-confidence SNP calls (*-m*>3), avoiding the formation of paralogs (*-M*<4), avoiding calling fixed alleles as separate loci (-n>0), and conforming to biological expectations (*--max_locus_stacks* = 2, as usually no more than 2 alleles are found in SNP markers of diploid organisms). *De novo* loci assemblies with selected parameters (*-m* 4 *-M* 3 *-n* 4 *--max_locus_stacks* 2) produced approximately 92±17 thousand unique sequence stacks per individual of *V. scotiaensis*, with a mean coverage depth of 24±6x using the restriction enzyme SgrAI (Table 3). The restriction enzyme SbfI was used with both *V. osheai* and *E. ohtai*. Approximately 28±3 thousand stacks per individual were produced for *V. osheai,* with a mean coverage depth of 41±13x (optimal assembly parameters: *-m* 4 *-M* 3 *-n* 3 *--max_locus_stacks* 2). Lastly, approximately 27±3

thousand stacks per individual were produced for *E. ohtai*, with a mean coverage depth of 73±28x (optimal assembly parameters: *-m* 5 *-M* 3 *-n* 3 *--max_locus_stacks* 2).

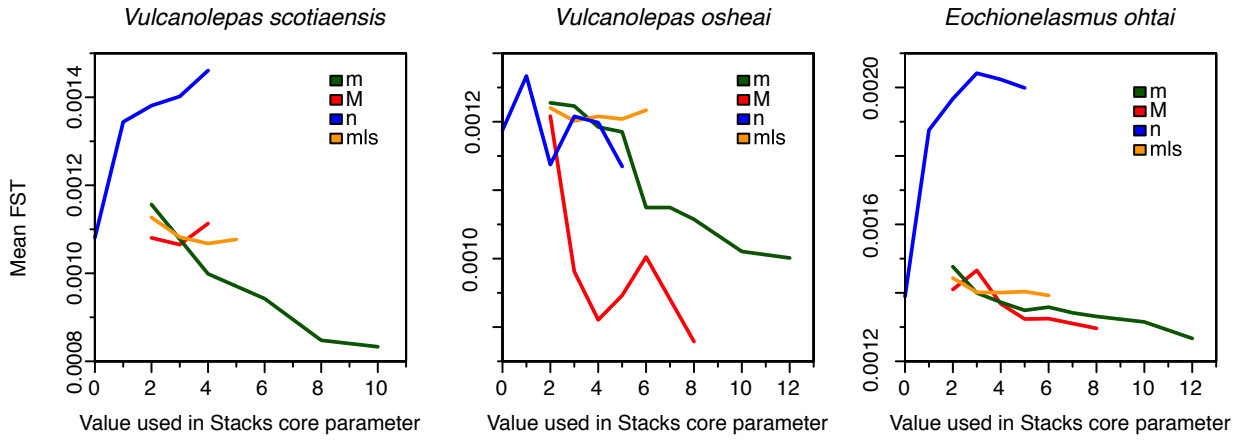


Figure 3. *De novo* loci assembly parameter influence on F_{ST} estimation for vent barnacle datasets. Species are indicated at the top of each subplot. Plots show the value of each one of the examined STACKS core parameters against the estimated mean F_{ST}. Line colors and legend indicate evaluated core parameters: the minimum number of identical raw reads required to create a stack of identical unique sequences for each individual (*-m* in green); the number of mismatches allowed between loci for each individual (*-M* in red); the number of mismatches allowed between loci when building the catalog of all loci in a species (*-n* in blue); and the maximum number of stacks at a single locus for each individual (*--max_locus_stacks*, abbreviated as mls, in yellow).

There were 33,507 loci (mean coverage depth of 20±6x) shared among 75% of individuals in all populations of *V. scotiaensis*, of which 28,270 contained one or more SNPs. Significantly fewer markers were obtained from *V. osheai,* with 4,384 loci (mean coverage depth of 34±11x), but a larger fraction, 4,205, contained one or more SNPs. Lastly, there were 9,966 loci for *E. ohtai* (mean coverage depth of 60±22x), of which 9,430 contained one or more SNPs.

Population demographic inferences

Summary population genetic statistics revealed virtually identical levels of genetic diversity among populations of each species – in terms of percentage of polymorphic sites, mean observed heterozygosity and mean nucleotide diversity (Table 4). Overall, *V. scotiaensis* had the greatest genetic diversity, followed by *V. osheai* and *E. ohtai.* These patterns were maintained when summary statistics were calculated from variant positions alone, and from variant and fixed positions combined.

Table 4. Population genetic statistics calculated only from variant positions, and from both variant and fixed positions. Values indicate means ± standard deviation

Variant positions

Population	Private alleles	Variant sites	% polym. sites	Major allele frequency	Observed heterozygosity	Nucleotide diversity (π)	F_{IS}
E2	21,327	55,411	64.63	0.9403±0.0949	0.0880±0.1311	0.0981±0.1257	0.0504±0.2356
E9	19,600	55,415	61.51	0.9411±0.0938	0.0823±0.1245	0.0973±0.1261	0.0667±0.2544
Healy	3,446	10,274	63.29	0.9419±0.0970	0.0671±0.1034	0.0935±0.1285	0.1263±0.3041
Tangaroa	3,772	10,274	66.46	0.9422±0.0949	0.0723±0.1082	0.0938±0.1257	0.1058±0.2851
Tahi Moana	4,977	23,177	50.83	0.9503±0.0922	0.0523±0.0938	0.0800±0.1241	0.1305±0.3071
Founalei South	4,849	23,167	50.09	0.9507±0.0917	0.0555±0.0980	0.0795±0.1241	0.1078±0.2846
Mata Ua	5,274	23,161	52.48	0.9504±0.0917	0.0523±0.0922	0.0798±0.1229	0.1336±0.3106

All positions (variant and fixed)

Population	Private alleles	Sites	% polym. sites	Major allele frequency	Observed heterozygosity	Nucleotide diversity (π)	F_{IS}
E2	21,327	1,728,859	2.07	0.9981±0.0200	0.0028±0.0283	0.0031±0.0283	0.0016±0.0436
E9	19,600	1,728,867	1.97	0.9981±0.0200	0.0026±0.0265	0.0031±0.0283	0.0021±0.0469
Healy	3,446	382,282	1.70	0.9984±0.0173	0.0018±0.0200	0.0025±0.0265	0.0034±0.0539
Tangaroa	3,772	382,279	1.79	0.9984±0.0173	0.0019±0.0224	0.0025±0.0265	0.0028±0.0500
Tahi Moana	4,977	853,175	1.38	0.9986±0.0173	0.0014±0.0173	0.0022±0.0245	0.0035±0.0548
Founalei South	4,849	853,168	1.36	0.9987±0.0173	0.0015±0.0173	0.0022±0.0245	0.0029±0.0500
Mata Ua	5,274	853,151	1.42	0.9987±0.0173	0.0014±0.0173	0.0022±0.0245	0.0036±0.0557

Mean pairwise F_{ST} values indicate low differentiation among populations within each species (0.0011 in *V. scotiaensis*; 0.0011 in *V. osheai*; and in *E. ohtai* 0.0018 between Tahi Moana and Founalei S., 0.0021 between Tahi Moana and Mata Ua, and 0.0017 between Founalei S. and Mata Ua). Globally, neither Neighbor-Joining similarity dendrograms nor principal component analyzes of SNP data produced clustering patterns of genetic variation consistent with population genetic structuring within species. In all cases, none of the axes in the principal component analyses explained more than 7% of the observed variance in the SNP data.

Minor allele frequency spectra show that a majority of the alleles in each population have low frequencies (Fig. 5), as expected for stable populations near mutation-drift equilibrium. F_{IS} distributions for all populations were centered on zero indicating random mating and lack of population-sub structuring (Fig. 6). A small fraction of F_{IS} values were positive, including noticeable clusters of values near 1, could be indicative of markers falling on non-recombinant genomic regions, such as sex-determining regions

(Catchen *et al.* 2013a). Both minor allele frequency spectra and F_{IS} distributions were remarkably similar among populations of each species.

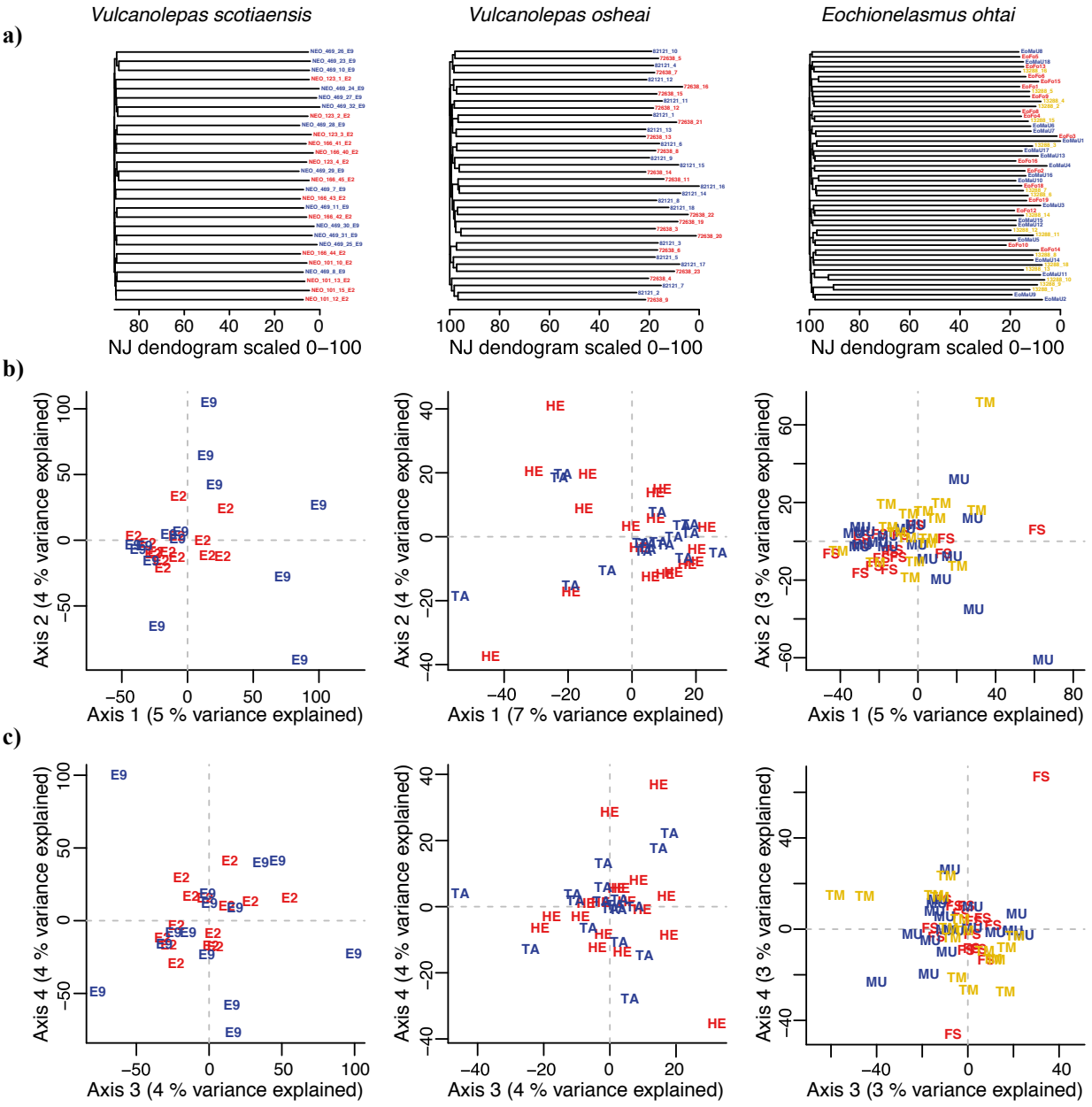


Figure 4. Genetic variability of SNP data from individuals for each vent barnacle species. Species are indicated in each column. **a)** Calibrated Neighbor-Joining dendrograms of genetic similarity from Euclidean distances among individuals per species. **b)** Plots of the first two axes of genetic variation among individuals per species found through principal component analyses. **c)** Plots of the third and fourth axes of genetic variation among individuals per species found through principal component analyses. Labels and colors indicate the source population per

species: *V. scotiaensis* E2 in red and E9 in blue; *V. osheai* Healy (HE) in red and Tangaroa (TA) in blue; *E. ohtai* Founalei South (FS) in red, Mata Ua (MU) in blue, and Tahi Moana (TM) in yellow.

DISCUSSION

Effects of de novo loci assembly parameter selection

Consistently with previous analyses by Mastretta-Yanes *et al.* (2014) and Catchen *et al.* (2013b), we found that the core parameters for *de novo* loci assembly in STACKS -m, -n, and -M, which limit the minimum number of identical raw reads required to create a stack of identical unique sequences for each individual, the maximum number of mismatches allowed between loci for each individual, and the maximum number of mismatches allowed between loci when building the catalog of all loci in a species, respectively, had the largest effects on the amount of loci and SNP generated from RAD sequence data. Although in this study we do not have access to a reference genome or sequencing replicates, with which parameter influences on *de novo* loci assembly error rates can be estimated, our analyses indicate that overall the population genetic inferences from datasets derived using a variety of assembly parameters were largely insensitive to parameter variations. This provides high confidence in the presented results, as the patterns of population genetic diversity here identified are most likely the result of true biological and ecological processes in the examined vent barnacle species, rather than methodological artifacts.

No evidence supporting seamount-driven isolation in vent barnacle populations

Altogether, we find no support for the seamount endemicity hypothesis in deep-sea hydrothermal-vent barnacles after examining thousands of genome-wide SNPs obtained from RAD-seq data. The patterns of genetic variation among individuals collected from seamount and spreading ridges, separated horizontally by hundreds of kilometers and vertically by hundreds of meters, did not reject the null hypothesis of panmixia within each species. Contrary to the predictions from the seamount endemicity hypothesis, we did not find higher population genetic structuring between seamount populations than between populations from spreading ridges.

Non-equilibrium processes, such as population expansion and recent colonization, can also lead to a lack of population structuring and apparently high gene flow. Evidence suggestive of non-equilibrium dynamics has been presented in populations of some of vent species, particularly in populations from fast-spreading ridges (Vrijenhoek 2010). Commonly invoked process to explain negative Tajima's D and Fu's F_S values and star-like haplotype networks in mitochondrial data, include bottlenecks and founder events resulting from processes such as catastrophic eruptions, vent formations and disappearances (Won *et al.*

2003; Hurtado *et al.* 2004; Plouviez *et al.* 2009; Teixeira *et al.* 2011; Thaler *et al.* 2011; Beedessee *et al.* 2013; Thaler *et al.* 2014). Nonetheless, minor allele frequency spectra for vent barnacle populations in this study were consistent with mutation-drift equilibrium, as a majority of the alleles in each population have low frequencies (Luikart *et al.* 1998), with a long tail of alleles at intermediate frequencies approaching zero proportions. Therefore, we conclude that the lack of population structuring and high gene flow among examined vent barnacle populations does not seem to be caused by non-equilibrium processes, but rather by sufficient genetic exchange among among populations.

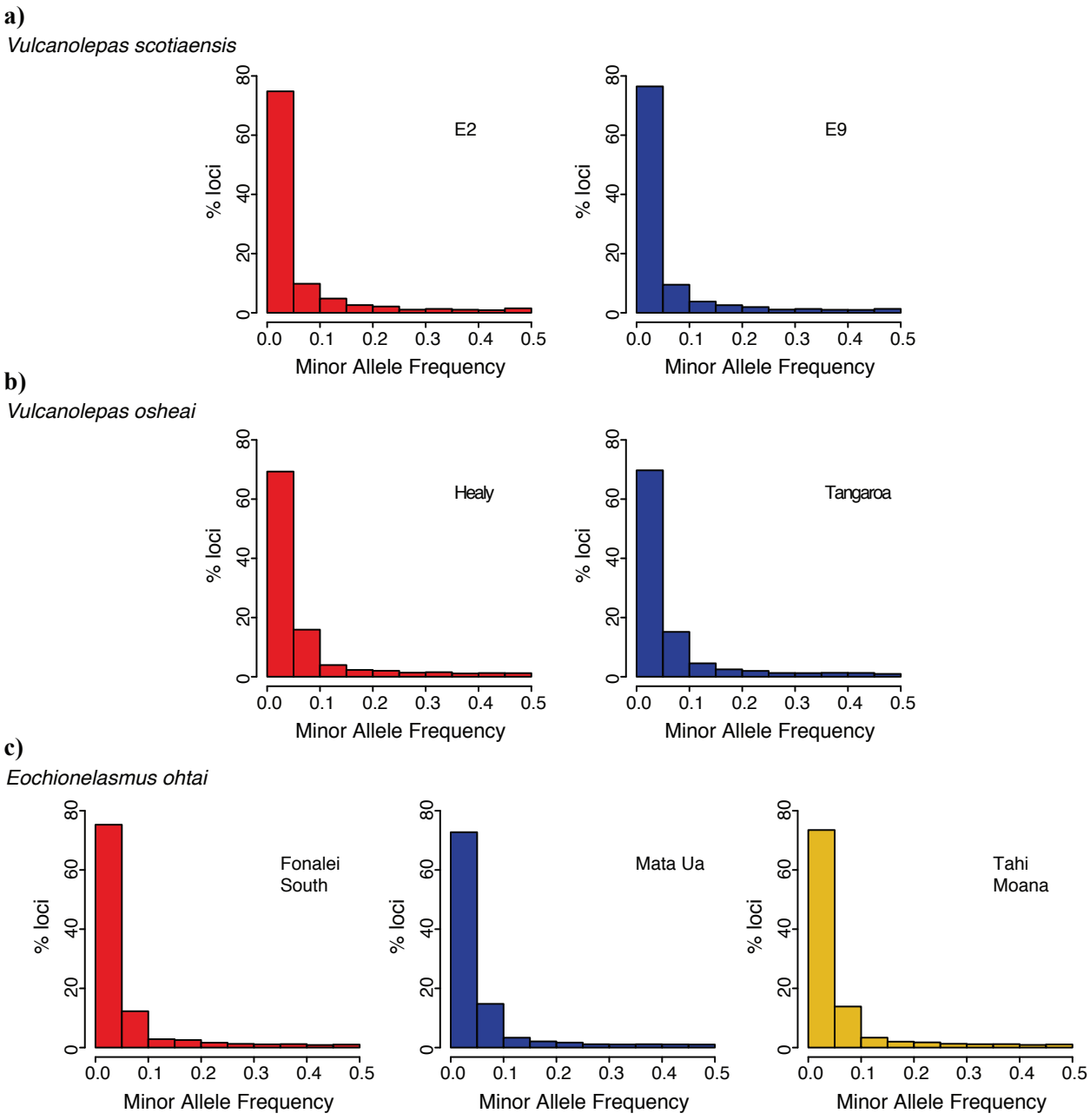


Figure 5. Minor allele frequency spectra from SNP loci for each vent barnacle population, per species. Labels and colors indicate the source population per species: **a)** *V. scotiaensis* – E2 in red and E9 in blue; **b)** *V. osheai* – Healy in red and Tangaroa in blue; **c)** *E. ohtai* – Founalei South in red, Mata Ua in blue, and Tahi Moana in yellow.

Larval development strategies have been long been hypothesized to be determinant of the dispersal potential of a species. Species in the vent barnacle clade A (*sensu* Herrera *et al.* (Chapter 3), to which *V. osheai* and *V. scotiaensis* belong, have large lecithotrophic naupliar larvae (Tunnicliffe & Southward 2004; Watanabe *et al.* 2004). *E. osheai* is known to produce large eggs (Newman *et al.* 2006), and thus presumably also has lecithotrophic larvae. It is often assumed that the duration in the water column for lecithotrophic larvae is limited by yolk reserves and thus lecithotrophic larvae are expected to have a shorter development time than planktotrophic larvae, and therefore smaller dispersal potential overall. However, significant evidence has accumulated against this hypothesis. For example, results from a colonization study in the East Pacific Rise suggest that the large lecithotrophic larvae can successfully disperse over long distances exceeding 300 kilometers (Mullineaux *et al.* 2010) (the horizontal distances among barnacle populations in this study range between *ca.* 100 and 400 kilometers). Additionally, temperatures in the deep ocean are known to reduce metabolic and development rates and therefore enhance dispersal (O'connor *et al.* 2007). Large yolk reserves, combined with cold sea-bottom temperatures are known to yield pelagic larval durations of up to 100 days in species in the vent barnacle clade A (Yorisue *et al.* 2013). Taken together, these characteristics provide a potential mechanism for enabling dispersal over long distances in vent barnacles and may partially account for the absence of population genetic structuring.

Local connectivity patterns in the East Scotia Ridge

Consistent with our observations of genetic variability in *V. scotianensis*, recent population genetic studies of other endemic species from hydrothermal vents in the East Scotia Ridge (ESR) –namely the yeti crab *Kiwa* sp. and two gastropod species – show patterns of no population differentiation along the ridge, between the E2 and E9 sites (Roterman 2013). Interestingly, one of the gastropod species, the limpet *Leptodrilus* sp., also occurs in the neighboring Kemp caldera in of the South Sandwich Arc, a region from which *V. scotiaensis* is also known (Rogers *et al.* 2012). Roterman (2013) found significant differentiation between the ridge and Kemp caldera limpet populations, which is separated from the ridge axis by only approximately 96 kilometers to the east, whereas the E2 and E9 sites are separated from each other by approximately 440 kilometers. Nonetheless E2 and E9 sites are a located at similar depths of approximately 2,500 meters, whereas the Kemp caldera vents are approximately 1,000 meters shallower, which could explain the observed differentiation. Testing for population differentiation in *V. scotiaensis*

between the ridge axis and the Kemp caldera would provide a good test for the generality of the observed depth isolation pattern in the limpet species.

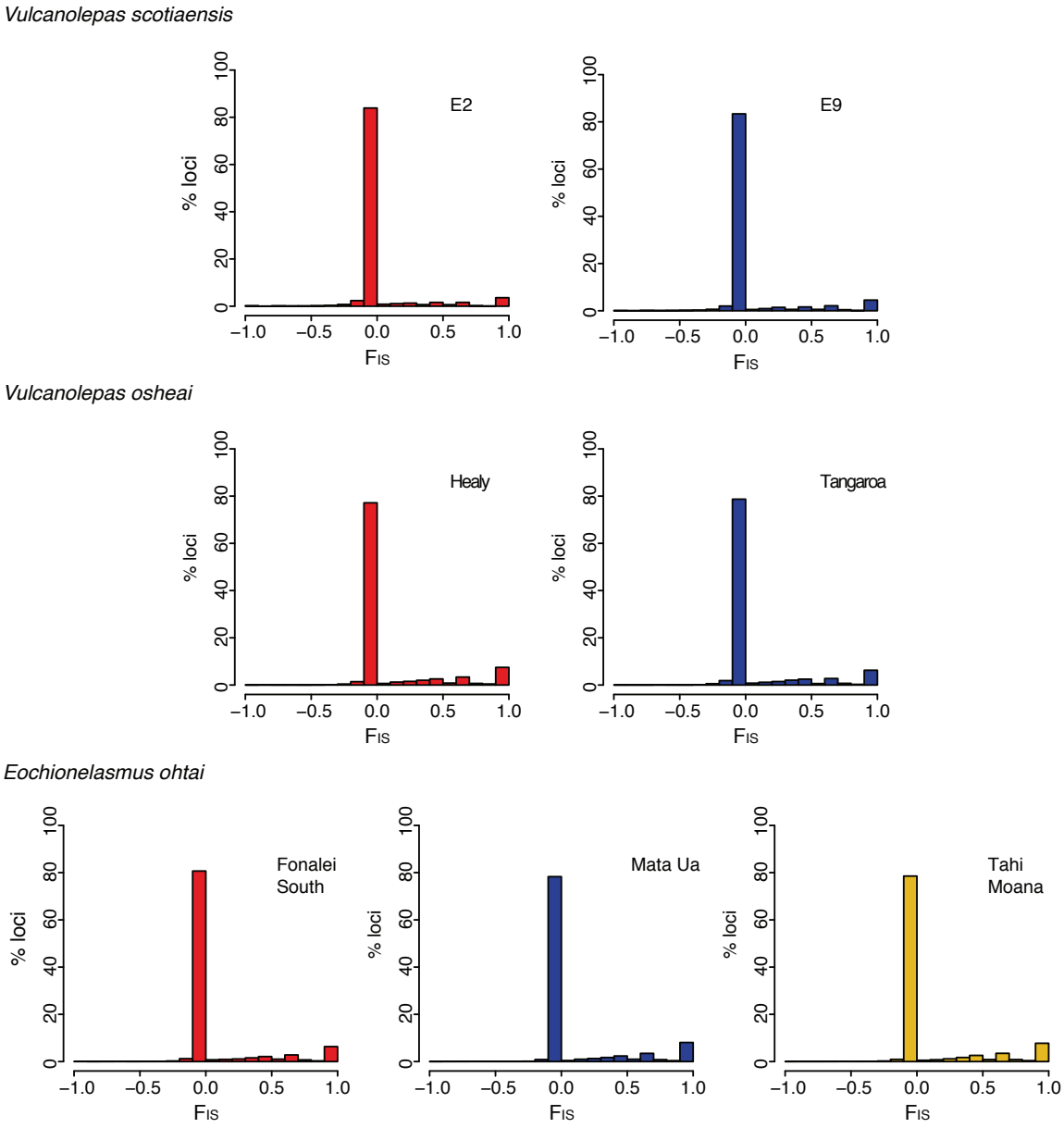


Figure 6. Frequency spectra of Wright's inbreeding index (F_{IS}) values from SNP loci for each vent barnacle population, per species. Labels and colors indicate the source population per species: **a)** *V. scotiaensis* – E2 in red and E9 in blue; **b)** *V. osheai* – Healy in red and Tangaroa in blue; **c)** *E. ohtai* –Founalei South in red, Mata Ua in blue, and Tahi Moana in yellow.

Although mean water fluxes in the ESR have an easterly or north-easterly direction, largely influenced by the Antarctic Circumpolar Current (Meredith *et al.* 2008), the local hydrography and flows on the E2, E9

and Kemp vent sites remain unknown. The ESR has an intermediate spreading rate of approximately 62 to 70 millimeters per year (Livermore 2003), which causes the formation of deep axial valleys with high walls that likely modify local current regimes. Mid-ocean ridges with similar characteristics are known to generate significant flows along and on the flanks of the ridge axes (Thurnherr *et al.* 2008; Thurnherr *et al.* 2011; Lavelle *et al.* 2012), thus potentially facilitating dispersal among spatially separated vent fields. Although hydrothermal vents at ridges with intermediate and low spreading rates – like the East Scotia Ridge – occur spatially less frequently along the ridge than vents at faster spreading systems – like the East Pacific Rise – (Baker & German 2004), they tend to be significantly more stable over time, in some cases for thousands to hundreds of thousands of years (Lalou *et al.* 1995; Jamieson *et al.* 2013). Therefore, local vent formations and extinctions in these systems may not be recurrent enough to significantly influence genetic variability patterns in vent fauna. Finally, potential undiscovered vent sites along the East Scotia Ridge axis (German *et al.* 2000) could act as stepping-stones for vent barnacles and other vent species, and thus facilitate dispersal and gene-flow.

Local connectivity patterns in the Kermadec arc

The only previous population genetic study in seamount populations from the Kermadec arc (Smith *et al.* 2004) found heterogeneity in allozyme allelic frequencies and significant genetic differentiation between populations of *Bathymodiolus* vent mussels from two active seamounts of the Kermadec arc, Rumble V and Rumble III, which were separated horizontally by approximately 50 km. However, this same study failed to reproduce the pattern of population structuring between seamount mussel populations when examining mitochondrial DNA sequences; therefore, it is possible that this pattern was an artifact of the allozyme markers. The use of other nuclear markers, such as SNP would clarify this discrepancy. Consistent with our results indicating high-gene flow among barnacle populations from Kermadec arc seamounts, Herrera *et al.* (Chapter 3) examined the phylogenetic relationships of vent barnacles and found no evidence of cryptic differentiation among *V. oseahi* individuals from Brothers, Clarck, Healy or Tangaroa seamounts.

Physical oceanographic processes could facilitate high non-directional dispersal and gene flow among seamounts in the southern Kermadec arc. Recent studies focusing on Brothers (Lavelle *et al.* 2008) and Rumble III seamounts (Stevens *et al.* 2014) indicate that circulation patterns in this region are highly dynamic over time, being largely influenced by isotropic tidal flows and mesoscale eddies, such as the East Cape Eddy. These studies indicate that flows speeds over these seamounts can range between 5 and 10 centimeters per second, which could potentially disperse larvae between 4 and 8 kilometers per day, and more than 130 to 260 kilometers in just a month. Futhermore, Stevens *et al.* (2014) found no evidence

of Taylor column formation on the Rumble III seamount that could cause larval or particle retention around the summits.

The Kermadec arc is a region of high volcanic activity (de Ronde *et al.* 2007), which likely leads to catastrophic local extinctions, and frequent formation and disappearance of vents. Just between 2006 and 2012 there were at least three reported eruptions in active seamounts (Dodge 2009; Watts *et al.* 2012; Jutzeler *et al.* 2014). However, active seamounts usually host multiple vent fields, which are likely not affected by volcanic and venting dynamics simultaneously or at the same rates. We hypothesize that presence of multiple populations on other active neighboring seamounts in the southern Kermadec arc, such as Brothers, Rumble III, and Clark (A. Rowden, M. Clark, and D. Bowden personal communication), combined with potential high dispersal and gene-flow rates, may jointly act as buffers that prevent significant losses of genetic diversity due to vent instability. Samples from additional populations and further comparative studies with multiple taxa would allow us to elucidate more general patterns for vent populations in these seamounts.

Local connectivity patterns in the Lau basin

Consistent with our results from Lau basin populations, Plouviez *et al.* (2013) found no signatures of genetic differentiation or potential population structuring at horizontal scales from hundreds of meters up to 50 kilometers in *Eochionelasmus ohtai* from the Manus Basin, using mitochondrial markers. Although populations from the Manus basin have been considered a potentially different sub-species from the one inhabiting the Lau and North Fiji basins (Newman *et al.* 2006), a recent phylogenetic study by Herrera *et al.* (Chapter 3) found no evidence of genetic differentiation between Manus and Lau basin individuals of *E. ohtai*. Similar signatures of panmixia have been found in shrimp (*Chorocaris* sp. 2) and gastropod snail (*Ifremeria nautilei*) vent species within the Manus and Lau basins (Thaler *et al.* 2011; Thaler *et al.* 2014). Contrastingly Thaler *et al.* (2014) found significant structuring in populations of the squat lobster *Munidopsis lauensis* within Manus basin, which the authors attributed to potential larvae behaviors and water flow directionality.

In addition to the populations included in this study, *E. ohtai* is known from vents in other active systems in the Lau Basin, including the Eastern Lau Spreading Center and the North-East Lau Spreading Center (Herrera *et al.* Chapter 3), which can serve as dispersal stepping-stones for dispersal throughout the basin. The Lau basin has an extremely diverse and active geologic setting, with multiple isolated back-arc spreading ridges, axial volcanoes and active arc and back-arc seamounts hosting hydrothermal vent communities. Volcanic eruptions that create and destroy hydrothermal vent communities are also

common in this region (Resing *et al.* 2011; Bohnenstiehl *et al.* 2014; Embley *et al.* 2014). However, as discussed earlier, we suggest that the presence of well-connected populations within a metapopulation in some vent species can act as buffers that prevent significant losses of genetic diversity due to vent instability. Ocean circulation at the scale of months in the Lau basin is dominated by isotropic flows caused by eddies, reaching speeds of up to 10 centimeters per second (Speer & Thurnherr 2012), which could help explain the lack of directionality or differentiation observed in the genetic pool of *E. otahi* and other vent species within the basin.

CONCLUSIONS

Comparative data from three species of barnacles from deep-sea hydrothermal vents revealed patterns of genetic variation inconsistent with the seamount endemicity hypothesis. These results reinforce the idea that within-basin structuring is rare while between-basin structuring is common, lending insight into the scale at which vent populations are structured. Both the Kermadec Arc and Lau Basin are areas of high interest for polymetallic sulphide mining (Boschen *et al.* 2013). Although these results indicate that populations of vent barnacles in the examined populations could be resilient to potential disturbance from local mining restricted to focal sites, we suggest that further multi-species and time-series studies in these and other target and neighboring populations should take place in order to better assess the potential impacts of this extractive activity. This study constitutes the first use of genome-wide SNP data to examine patterns of population genetic structuring and connectivity among populations of vent species. We suggest that the seamount endemicity hypothesis warrants further testing using high-resolution SNP data in other vent organisms with differing life history strategies (e.g., brooders) that may limit their dispersal potential, as well as in non-vent organisms, which are not exposed to evolutionary pressures imposed by the dynamic nature of hydrothermal vent systems.

ACKNOWLEDGEMENTS

This research was supported by the Office of Ocean Exploration and Research of the National Oceanic and Atmospheric Administration (NA09OAR4320129 to TS); the Division of Ocean Sciences of the National Science Foundation (OCE-1131620 to TS); the Division of Polar Programs of the National Science Foundation (PLR-0739675 to TS); the Astrobiology Science and Technology for Exploring Planets program of the National Aeronautics and Space Administration (NNX09AB76G to TS); and the Academic Programs Office (Ocean Ventures Fund to SH), the Ocean Exploration Institute (Fellowship

support to TMS) and the Ocean Life Institute of the Woods Hole Oceanographic Institution (internal grant to TMS and SH). For enabling access to key specimens we thank P. Tyler (NOCS), J. Copley (NOCS), A. Rogers (Oxford), N. Roterman (Oxford), M. Clark (NIWA), A. Rowden (NIWA), K. Schnabel (NIWA), S. Mills (NIWA), J. Resing (NOAA-PMEL), R. Embley (NOAA-PMEL), A. Reysenbach (PSU), M.K. Tivey (WHOI), the masters, crew, scientific personnel, and funding agencies of expeditions JC042, RR1211, TAN1104, TAN1206, and TN236. Specimens provided by the National Institute of Water and Atmospheric Research (NIWA) were collected under research programs: Ocean Survey 20/20 funded by Land Information New Zealand; and the Impact of resource use on vulnerable deep-sea communities (CO1X0906), funded by MBIE. We also thank E. Bors, A. Tarrant, and A. Reitzel for providing helpful comments that improved this manuscript.

REFERENCES

Alberto F, Raimondi PT, Reed DC*, et al.* (2011) Isolation by oceanographic distance explains genetic structure for *Macrocystis pyrifera* in the Santa Barbara Channel. *Molecular Ecology* **20**, 2543-2554.

Baco AR, Cairns SD (2012) Comparing molecular variation to morphological species designations in the deep-sea coral *Narella* reveals new Insights into seamount coral ranges. *PLoS One* **7**, e45555.

Baird NA, Etter PD, Atwood TS*, et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**, e3376.

Baker ET, German CR (2004) On the Global Distribution of Hydrothermal Vent Fields. In: *Mid-Ocean Ridges*, pp. 245-266. American Geophysical Union.

Beedessee G, Watanabe H, Ogura T*, et al.* (2013) High connectivity of animal populations in deep-sea hydrothermal vent fields in the Central Indian Ridge relevant to its geological setting. *PLoS One* **8**, e81570.

Beinart RA, Sanders JG, Faure B*, et al.* (2012) Evidence for the role of endosymbionts in regional-scale habitat partitioning by hydrothermal vent symbioses. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E3241-E3250.

Bohnenstiehl DR, Dziak RP, Matsumoto H, Conder JA (2014) Acoustic response of submarine volcanoes in the Tofua Arc and northern Lau Basin to two great earthquakes. *Geophysical Journal International* **196**, 1657-1675.

Boschen RE, Rowden AA, Clark MR, Gardner JPA (2013) Mining of deep-sea seafloor massive sulfides: A review of the deposits, their benthic communities, impacts from mining, regulatory frameworks and management strategies. *Ocean and Coastal Management* **84**, 54-67.

Catchen J, Bassham S, Wilson T*, et al.* (2013a) The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. *Molecular Ecology* **22**, 2864-2883.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013b) Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**, 3124-3140.

Cho W, Shank TM (2010) Incongruent patterns of genetic connectivity among four ophiuroid species with differing coral host specificity on North Atlantic seamounts. *Marine Ecology* **31**, 121-143.

Christensen NL, Bartuska AM, Brown JH*, et al.* (1996) The report of the Ecological Society of America committee on the scientific basis for ecosystem management. *Ecological Applications* **6**, 665-691.

Clague GE, Jones WJ, Paduan JB, Clague DA, Vrijenhoek RC (2012) Phylogeography of *Acesta* clams from submarine seamounts and escarpments along the western margin of North America. *Marine Ecology-an Evolutionary Perspective* **33**, 75-87.

D'Aloia CC, Bogdanowicz SM, Harrison RG, Buston PM (2014) Seascape continuity plays an important role in determining patterns of spatial genetic structure in a coral reef fish. *Molecular Ecology* **23**, 2902-2913.

de Forges BR, Koslow JA, Poore GC (2000) Diversity and endemism of the benthic seamount fauna in the southwest Pacific. *Nature* **405**, 944-947.

de Ronde CEJ, Baker ET, Massoth GJ*, et al.* (2007) Submarine hydrothermal activity along the mid-Kermadec Arc, New Zealand: Large-scale effects on venting. *Geochemistry, Geophysics, Geosystems* **8**, Q07007.

Dodge E (2009) *Catastrophic volcanic activity at Rumble III volcano based on EM300 bathymetry and direct seafloor imaging*, University of Washington.

Embley RW, Merle SG, Baker ET*, et al.* (2014) Eruptive modes and hiatus of volcanism at West Mata seamount, NE Lau basin: 1996–2012. *Geochemistry, Geophysics, Geosystems* **15**, 4093-4115.

Emerson KJ, Merz CR, Catchen JM*, et al.* (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 16196-16200.

German CR, Livermore RA, Baker ET*, et al.* (2000) Hydrothermal plumes above the East Scotia Ridge: an isolated high-latitude back-arc spreading centre. *Earth and Planetary Science Letters* **184**, 241-250.

Herrera S, Reyes-Herrera PH, Shank TM (Chapter 2) Genome-wide predictability of restriction sites across the eukaryotic tree of life.

Herrera S, Shank TM (Chapter 5) RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa.

Herrera S, Watanabe H, Shank T (Chapter 3) Evolutionary and biogeographical patterns of barnacles from deep-sea hydrothermal vents.

Hohenlohe PA, Bassham S, Etter PD*, et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* **6**, e1000862.

Hurtado LA, Lutz RA, Vrijenhoek RC (2004) Distinct patterns of genetic differentiation among annelids of eastern Pacific hydrothermal vents. *Molecular Ecology* **13**, 2603-2615.

Jamieson JW, Hannington MD, Clague DA*, et al.* (2013) Sulfide geochronology along the Endeavour Segment of the Juan de Fuca Ridge. *Geochemistry, Geophysics, Geosystems* **14**, 2084-2099.

Jutzeler M, Marsh R, Carey RJ*, et al.* (2014) On the fate of pumice rafts formed during the 2012 Havre submarine eruption. *Nature Communications* **5**, Artn 3660.

Kelly RP, Palumbi SR (2010) Genetic structure among 50 species of the northeastern Pacific rocky intertidal community. *PLoS One* **5**, e8594.

Lalou C, Reyss J-L, Brichet E, Rona PA, Thompson G (1995) Hydrothermal activity on a 105-year scale at a slow-spreading ridge, TAG hydrothermal field, Mid-Atlantic Ridge 26°N. *Journal of Geophysical Research: Solid Earth* **100**, 17855-17862.

Lavelle JW, Massoth GJ, Baker ET, de Ronde CEJ (2008) Ocean current and temperature time series at Brothers volcano. *Journal of Geophysical Research-Oceans* **113**, C09018.

Lavelle JW, Thurnherr AM, Mullineaux LS, McGillicuddy DJ, Ledwell JR (2012) The prediction, verification, and significance of flank jets at mid-ocean ridges. *Oceanography* **25**, 277-283.

Livermore R (2003) Back-arc spreading and mantle flow in the East Scotia Sea. *Geological Society, London, Special Publications* **219**, 315-331.

Luikart G, Allendorf FW, Cornuet JM, Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity* **89**, 238-247.

Mastretta-Yanes A, Arrigo N, Alvarez N*, et al.* (2014) Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources* **15**, 28–41.

Meredith MP, Garabato ACN, Gordon AL, Johnson GC (2008) Evolution of the deep and bottom waters of the Scotia Sea, Southern Ocean, during 1995-2005. *Journal of Climate* **21**, 3327-3343.

Metaxas A (2011) Spatial patterns of larval abundance at hydrothermal vents on seamounts: evidence for recruitment limitation. *Marine Ecology Progress Series* **437**, 103-117.

Mullineaux LS, Adams DK, Mills SW, Beaulieu SE (2010) Larvae from afar colonize deep-sea hydrothermal vents after a catastrophic eruption. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 7829-7834.

Mullineaux LS, Mills SW (1997) A test of the larval retention hypothesis in seamount-generated flows. *Deep-Sea Research Part I-Oceanographic Research Papers* **44**, 745-&.

Newman WA, Yamaguchi T, Southward AJ, Segonzac M (2006) Arthropoda, Crustacea, Cirripedia. In: *Handbook of Deep-Sea Hydrothermal Vent Fauna* (eds. Desbruyeres D, Segonzac M, Bright M), pp. 356-357. Denisia.

O'connor MI, Bruno JF, Gaines SD*, et al.* (2007) Temperature control of larval dispersal and the implications for marine ecology, evolution, and conservation. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 1266-1271.

Plouviez S, Schultz TF, McGinnis G*, et al.* (2013) Genetic diversity of hydrothermal-vent barnacles in Manus Basin. *Deep-Sea Research Part I-Oceanographic Research Papers* **82**, 73-79.

Plouviez S, Shank TM, Faure B*, et al.* (2009) Comparative phylogeography among hydrothermal vent species along the East Pacific Rise reveals vicariant processes and population expansion in the South. *Molecular Ecology* **18**, 3903-3917.

Prada C, Hellberg ME (2013) Long prereproductive selection and divergence by depth in a Caribbean candelabrum coral. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 3961-3966.

Reitzel AM, Herrera S, Layden MJ, Martindale MQ, Shank TM (2013) Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular Ecology* **22**, 2953-2970.

Resing JA, Rubin KH, Embley RW*, et al.* (2011) Active submarine eruption of boninite in the northeastern Lau Basin. *Nature Geoscience* **4**, 799-806.

Rogers AD, Morley S, Fitzcharles E, Jarvis K, Belchier M (2006) Genetic structure of Patagonian toothfish (*Dissostichus eleginoides*) populations on the Patagonian Shelf and Atlantic and western Indian Ocean Sectors of the Southern Ocean. *Marine Biology* **149**, 915-924.

Rogers AD, Tyler PA, Connelly DP*, et al.* (2012) The discovery of new deep-sea hydrothermal vent communities in the southern ocean and implications for biogeography. *PLoS Biology* **10**, e1001234.

Roterman CN (2013) *The Evolution and Population Genetics of Hydrothermal Vent Megafauna from the Scotia Sea*, University of Oxford.

Ryan WBF, Carbotte SM, Coplan JO*, et al.* (2009) Global multi-resolution topography synthesis. *Geochemistry, Geophysics, Geosystems* **10**, Q03014.

Smith PJ, McVeagh SM, Won Y, Vrijenhoek RC (2004) Genetic heterogeneity among New Zealand species of hydrothermal vent mussels (Mytilidae : *Bathymodiolus*). *Marine Biology* **144**, 537-545.

Speer K, Thurnherr AM (2012) The Lau Basin float experiment (LAUB-FLEX). *Oceanography* **25**, 284-285.

Stevens CL, Consalvey M, Devine JA, Clark MR (2014) Mixing and transport near the shallow-crested Rumble III seamount and the implications for plankton distribution. *New Zealand Journal of Marine and Freshwater Research* **48**, 194-215.

Teixeira S, Cambon-Bonavita MA, Serrao EA, Desbruyeres D, Arnaud-Haond S (2011) Recent population expansion and connectivity in the hydrothermal shrimp *Rimicaris exoculata* along the Mid-Atlantic Ridge. *Journal of Biogeography* **38**, 564-574.

Thaler AD, Plouviez S, Saleu W*, et al.* (2014) Comparative population structure of two deep-sea hydrothermal-vent-associated decapods (*Chorocaris* sp 2 and *Munidopsis lauensis*) from southwestern Pacific back-arc basins. *PLoS One* **9**, e101345.

Thaler AD, Zelnio K, Saleu W*, et al.* (2011) The spatial scale of genetic subdivision in populations of *Ifremeria nautilei,* a hydrothermal-vent gastropod from the southwest Pacific. *BMC Evolutionary Biology* **11**, 372.

Thoma J, Pante E, Brugler M, France S (2009) Deep-sea octocorals and antipatharians show no evidence of seamount-scale endemism in the NW Atlantic. *Marine Ecology Progress Series* **397**, 25-35.

Thurnherr AM, Ledwell JR, Lavelle JW, Mullineaux LS (2011) Hydrography and circulation near the crest of the East Pacific Rise between 90 degrees and 10 degrees N. *Deep-Sea Research Part I-Oceanographic Research Papers* **58**, 365-376.

Thurnherr AM, Reverdin G, Bouruet-Aubertot P*, et al.* (2008) Hydrography and flow in the Lucky Strike segment of the Mid-Atlantic Ridge. *Journal of Marine Research* **66**, 347-372.

Tunnicliffe V, Koop BF, Tyler J, So S (2010) Flatfish at seamount hydrothermal vents show strong genetic divergence between volcanic arcs. *Marine Ecology-an Evolutionary Perspective* **31**, 158-167.

Tunnicliffe V, Southward AJ (2004) Growth and breeding of a primitive stalked barnacle *Leucolepas longa* (Cirripedia : Scalpellomorpha : Eolepadidae : Neolepadinae) inhabiting a volcanic seamount off Papua New Guinea. *Journal of the Marine Biological Association of the United Kingdom* **84**, 121-132.

Van Dover CL (2000) *The Ecology of Deep-Sea Hydrothermal Vents* Princeton University Press, Princeton, N.J.

Van Dover CL (2010) Mining seafloor massive sulphides and biodiversity: what is at risk? *ICES Journal of Marine Science* **68**, 341-348.

Vrijenhoek RC (2010) Genetic diversity and connectivity of deep-sea hydrothermal vent metapopulations. *Molecular Ecology* **19**, 4391-4411.

Watanabe H, Kado R, Tsuchida S*, et al.* (2004) Larval development and intermoult period of the hydrothermal vent barnacle *Neoverruca* sp. *Journal of the Marine Biological Association of the United Kingdom* **84**, 743-745.

Watanabe H, Tsuchida S, Fujikura K*, et al.* (2005) Population history associated with hydrothermal vent activity inferred from genetic structure of neoverrucid barnacles around Japan. *Marine Ecology-Progress Series* **288**, 233-240.

Watts AB, Peirce C, Grevemeyer I*, et al.* (2012) Rapid rates of growth and collapse of Monowai submarine volcano in the Kermadec Arc. *Nature Geoscience* **5**, 510-515.

White C, Selkoe KA, Watson J*, et al.* (2010) Ocean currents help explain population genetic structure. *Proceedings of the Royal Society B-Biological Sciences* **277**, 1685-1694.

Wilson RR, Kaufmann RS (1986) Seamount biota and biogeography. In: *Seamounts, Islands, and Atolls* (eds. Keating BH, Fryer P, Batiza R, Boehlert GW), Washington DC.

Won Y, Young CR, Lutz RA, Vrijenhoek RC (2003) Dispersal barriers and isolation among deep-sea mussel populations (Mytilidae : *Bathymodiolus*) from eastern Pacific hydrothermal vents. *Molecular Ecology* **12**, 169-184.

Yorisue T, Kado R, Watanabe H*, et al.* (2013) Influence of water temperature on the larval development of *Neoverruca* sp and *Ashinkailepas seepiophila*-Implications for larval dispersal and settlement in the vent and seep environments. *Deep-Sea Research Part I-Oceanographic Research Papers* **71**, 33-37.

SUPPORTING INFORMATION

Table S1. Predictions of number of RAD-tags in thoraciacan barnacles using SbfI. Data for *Daphnia pulex* obtained from the U.S. National Center for Biotechnology Information (NCBI) WGS database. Observed frequency of recognition sequences and calculated probability based on a trinucleotide genome composition model were generated following the methodology described by Herrera *et al.*(Herrera *et al.* 2014). Data for known barnacle genome sizes obtained from the Animal Genome Size Database (http://ww.genomesize.com). C-value is the amount of DNA in picograms in the nucleous, where the genome size in Mbp = 978 x C-value

Species	**Common name**	**C-value**	**Genome size (Mbp)**	**Observed frequency of SbfI recognition sites per bp**	**Probability of SbfI recognition site per bp**
Genebank WGS					
Daphnia pulex	Water flea		158.61	7.48E-06	5.90E-06

Species	**Common name**	**C-value**	**Genome size (Mbp)**	**Predicted number of SbfI recognition sites based on *D. pulex* observed frequency**	**Predicted number of SbfI RAD-tags based on *D. pulex* observed frequency**	**Predicted number of SbfI recognition sites based on *D. pulex* trinucleotide genome composition probability**	**Predicted number of SbfI RAD-tags based on *D. pulex* trinucleotide genome composition probability**
Animal Genome Size Database							
Balanus amphitrite	Striped barnacle	0.74	723.72	5,411.68	10,823.35	4,267.22	8,534.43
Balanus amphitrite	Striped barnacle	1.4	1,369.20	10,238.31	20,476.61	8,073.11	16,146.22
Balanus cariosus	Thatched barnacle	1.4	1,369.20	10,238.31	20,476.61	8,073.11	16,146.22
Balanus eburneus	Ivory barnacle	1.26	1,232.28	9,214.48	18,428.95	7,265.80	14,531.60
Chthalamus sp.	Acorn barnacle	1.23	1,202.94	8,995.08	17,990.17	7,092.80	14,185.61
Unknown sp.	Goose barnacle	1.46	1,427.88	10,677.09	21,354.18	8,419.10	16,838.20
Mitella polymerus	Pacific goose barnacle	0.9	880.20	6,581.77	13,163.54	5,189.86	10,379.71
Tetraclita rubescens	Volcano barnacle	2.6	2,542.80	19,014.00	38,027.99	14,992.92	29,985.84

CHAPTER 5

RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa

ABSTRACT

Species delimitation is problematic in many taxa due to the difficulty of evaluating predictions from species delimitation hypotheses, which chiefly relay on subjective interpretations of morphological observations and/or DNA sequence data. This problem is exacerbated in recalcitrant taxa for which genetic resources are scarce and inadequate to resolve questions regarding evolutionary relationships and uniqueness. In this case study we demonstrate the empirical utility of restriction site associated DNA sequencing (RAD-seq) by unambiguously resolving phylogenetic relationships among recalcitrant octocoral taxa with divergences greater than 80 million years. We objectively infer robust species boundaries in the genus *Paragorgia,* which contains some of the most important ecosystem engineers in the deep-sea, by testing alternative taxonomy-guided or unguided species delimitation hypotheses using the Bayes factors delimitation method (BFD*) with genome-wide single nucleotide polymorphism data. We present conclusive evidence rejecting the current morphological species delimitation model for the genus *Paragorgia* and indicating the presence of cryptic species boundaries associated with environmental variables. We argue that the suitability limits of RAD-seq for phylogenetic inferences in divergent taxa cannot be assessed in terms of absolute time, but depend on taxon-specific factors such as mutation rate, generation time and effective population size. We show that classic morphological taxonomy can greatly benefit from integrative approaches that provide objective tests to species delimitation hypothesis. Our results pave the way for addressing further questions in biogeography, species ranges, community ecology, population dynamics, conservation, and evolution in octocorals and other marine taxa.

INTRODUCTION

Species delimitation is problematic in many taxa due to the difficulty of evaluating predictions from species delimitation hypotheses derived using different species concepts. Species concepts set particular expectations of the properties used to support species delimitations (De Queiroz 2007). For example, the classic biological species concept requires intrinsic reproductive isolation between heterospecific organisms and interbreeding among homospecific organisms resulting in viable and fertile descendants (Mayr 1942; Dobzhansky 1970). In many cases, if not the majority, it is difficult to evaluate behavioral, reproductive, and ecological properties due to technical limitations of field or laboratory work, which largely determine the kind of observations and data that can be obtained. In these cases researchers conventionally rely on morphological observations and/or DNA sequence data to generate species delimitation hypotheses.

Although there have been significant attempts at developing statistical methods to objectively identify species-diagnostic morphological discontinuities (e.g., Zapata & Jimenez 2012), most species delimitations continue to be performed subjectively based on assessments made by specialized taxonomists. Molecular phylogenetic analyses of DNA sequences provide an independent way to test these species delimitation hypotheses utilizing a variety of methods, ranging from variability thresholds of barcode sequences (Hebert *et al.* 2003), to probabilistic coalescent-based model methods (Pons *et al.* 2006; Yang & Rannala 2010; Fujisawa & Barraclough 2013; Grummer *et al.* 2014). These molecular methods rely on informative DNA sequence markers, and in many cases on resolved phylogenies.

The sub-class Octocorallia (Phylum Cnidaria), which includes animals known as gorgonians, sea pens, and soft corals, is an example of a recalcitrant group where species delimitations are problematic. Octocorals are predominantly a deep-sea group (Cairns 2007; Roberts & Cairns 2014) and therefore are extremely difficult to observe and collect. Classic morphology-based species delimitation and identification in this group is arduous for non-specialists, and challenging to replicate among taxonomists (Daly *et al.* 2007; McFadden *et al.* 2010b). Variations in octocoral colony architecture and micro-skeletal structures – sclerites – are used as species diagnostic characters (Bayer 1956). However, studies over the last 15 years have shown that in many cases species delimitations and systematics based on these morphological traits keep little to no correspondence with the patterns of genetic diversity and relatedness inferred using mitochondrial and ribosomal DNA sequence markers (McFadden *et al.* 2006; Clark *et al.* 2007; France 2007; Dueñas & Sánchez 2009). A confounding factor when analyzing mitochondrial DNA markers is the fact that anthozoans, including octocorals, have slow rates of sequence evolution relative to

other metazoans (Shearer *et al.* 2002; Hellberg 2006). Furthermore, octocoral mitochondrion is unique among eukaryotes by having a functional DNA mismatch repair gene — *mtMutS* — which presumably is responsible for the extremely low sequence variability observed in this group (Bilewitch & Degnan 2011). Traditional molecular markers have thus been remarkably insufficient to resolve relationships at all taxonomic levels within the octocorals (Berntson *et al.* 2001; France *et al.* 2002; Mcfadden *et al.* 2004; Smith *et al.* 2004; Thoma *et al.* 2009; Dueñas *et al.* 2014). Alternative nuclear markers, such as the ITS2 and *SRP54* have been used to examine interspecific and intraspecific relationships (Aguilar & Sánchez 2007; Concepcion *et al.* 2007; Grajales *et al.* 2007; Herrera *et al.* 2010); however, their application and impact has been limited due to issues regarding intragenomic variability (Sanchez & Dorado 2008) and low sequencing reliability (Mcfadden *et al.* 2010a). These long-standing technical problems have caused fundamental questions in octocorals regarding species differentiation, systematics, diversity, biogeography, and species ranges to remain unanswered.

Technological developments in next-generation sequencing platforms and library preparation methodologies have made genomic resources increasingly accessible and available for the study of non-model organisms, thus offering a great opportunity to overcome the difficulties inherent to the use of traditional sequencing approaches. One of these methodologies is restriction-site-associated DNA sequencing (RAD-seq), which combines enzymatic fragmentation of genomic DNA with high-throughput sequencing for the generation of large numbers of markers (Baird *et al.* 2008). RAD-seq has shown great promise to resolve difficult phylogenetic, phylogeographic, and species delimitation questions in diverse groups of eukaryotes (Emerson *et al.* 2010; Nadeau *et al.* 2012; Wagner *et al.* 2012; Eaton & Ree 2013; Jones *et al.* 2013; Cruaud *et al.* 2014; Escudero *et al.* 2014; Hipp *et al.* 2014; Leache *et al.* 2014; Herrera *et al.* Chapter 3), including cnidarians (Reitzel *et al.* 2013) and most recently deep-sea octocorals (Pante *et al.* 2014). The number of orthologous restriction sites that can be retained across taxa, which decreases as divergence increases, limits the usefulness of RAD-seq for these kinds of studies. *In silico* studies in model organisms indicate that RAD-seq can be used to infer phylogenetic relationships in young groups of species (up to 60 million years old), such as *Drosophila* (Rubin *et al.* 2012; Cariou *et al.* 2013; Seetharam & Stuart 2013)*;* however, the real limits of this technique have not been significantly explored.

In this study we aim to empirically explore the limits of RAD-seq to solve questions in phylogenetics and species delimitation. We focus on the recalcitrant *Anthomastus-Corallium* clade of octocorals (sensu McFadden *et al.* 2006) to test the utility of RAD-seq to resolve phylogenetic relationships among divergent taxa, and to infer objective species boundaries. Corals in the *Anthomastus-Corallium* clade (hereafter referred as the AC clade) are among the most conspicuous, widely distributed, and ecologically

important benthic invertebrates in deep-water ecosystems (Roberts *et al.* 2009; Wating *et al.* 2011). This clade is constituted by more than 100 species defined morphologically, divided in 10 genera, and three families (World Register of Marine Species at http://www.marinespecies.org accessed on 2014-10-10), spanning a divergence time of over 100 million years (Ardila *et al.* 2012; Herrera *et al.* 2012). However, species delimitations and phylogenetic relationships in this clade, as in other octocorals, are controversial and conflictive (Herrera *et al.* 2010; Ardila *et al.* 2012; Herrera *et al.* 2012). Many of the species in this group are considered species indicators of Vulnerable Marine Ecosystems (e.g. ICES 2013), with some of them considered endangered (CITES 2014). Accurate species identifications, as well as complete inventories and knowledge of species ranges, are therefore critical to ensure the effectiveness and appropriateness of conservation and management policies.

RESULTS

Morphological species identifications

Using current species descriptions, colony observations, and scanning electron microscopy of sclerites, we identified a total of 12 putative morphological species among the 44 examined specimens from the AC clade (Table S1). These species correspond to the genera *Paragorgia* (*P. arborea, P. stephencairnsi, P. johnsoni, P. maunga, P. alisonae, P. kaupeka,* and *P. coralloides*) and *Sibogagorgia* (*S. cauliflora*) of the family Paragorgiidae; *Hemicorallium* (*H. laauense-imperiale*) and *Corallium* of the family Coralliidae; and *Anthomastus* and *Heteropolypus* of the family Alcyoniidae.

Octocorals are amenable to RAD sequencing

We generated a dense genome-wide set of genetic markers from the 44 AC clade specimens via RAD sequencing, using the 6-cutter restriction enzyme PstI, and used them to perform phylogenetic inferences and species delimitation analyses. We obtained roughly 3.9 ± 1.4 million reads (average ± standard deviation) per individual, of which 74.3 ± 8.1% were retained after stringent quality filtering steps (Table S2).

Optimization of RAD-loci clustering parameters

To examine the sensitivity of the phylogenetic inference to the clustering parameters used to identify loci and create nucleotide matrices in the program pyRAD (Eaton 2014), we investigated different combinations of clustering thresholds (c 0.80, 0.85 and 0.90) and minimum number of taxa per locus (m 4, 6, and 9) in a reduced **'backbone'** matrix (hereafter matrix names will be highlighted in bold) containing one individual from each of the 12 morphological species. The 9 resulting **backbone** matrices

ranged in the total number of loci per matrix from approximately 9 to 60 thousand loci, increasing dramatically as the minimum number of taxa per locus was reduced (Table S3). In contrast, the different clustering thresholds did not have a significant effect on the total number of loci, but rather on the number of variable sites and, most importantly, on the number of phylogenetically informative sites (Table S3). Each resulting **backbone** matrix analyzed in RAxML (Stamatakis 2006) produced identical strongly-supported tree topologies (Fig. S1). We selected c 0.80 (80% similarity among sequences) and m 9 (minimum coverage of taxa per locus of 75%) as the optimal combination of loci-clustering parameters because they minimized the proportion of missing data (0.20) in the matrix while maximizing the fraction of variable sites that were phylogenetically informative (0.24) (Table S3). The proportion of shared loci among individuals of Paragorgiidae and Coralliidae, lineages whose split has been estimated to be between 80-150 million years ago (Ardila *et al.* 2012; Herrera *et al.* 2012), was remarkably high (70-80%) (Fig. 1).

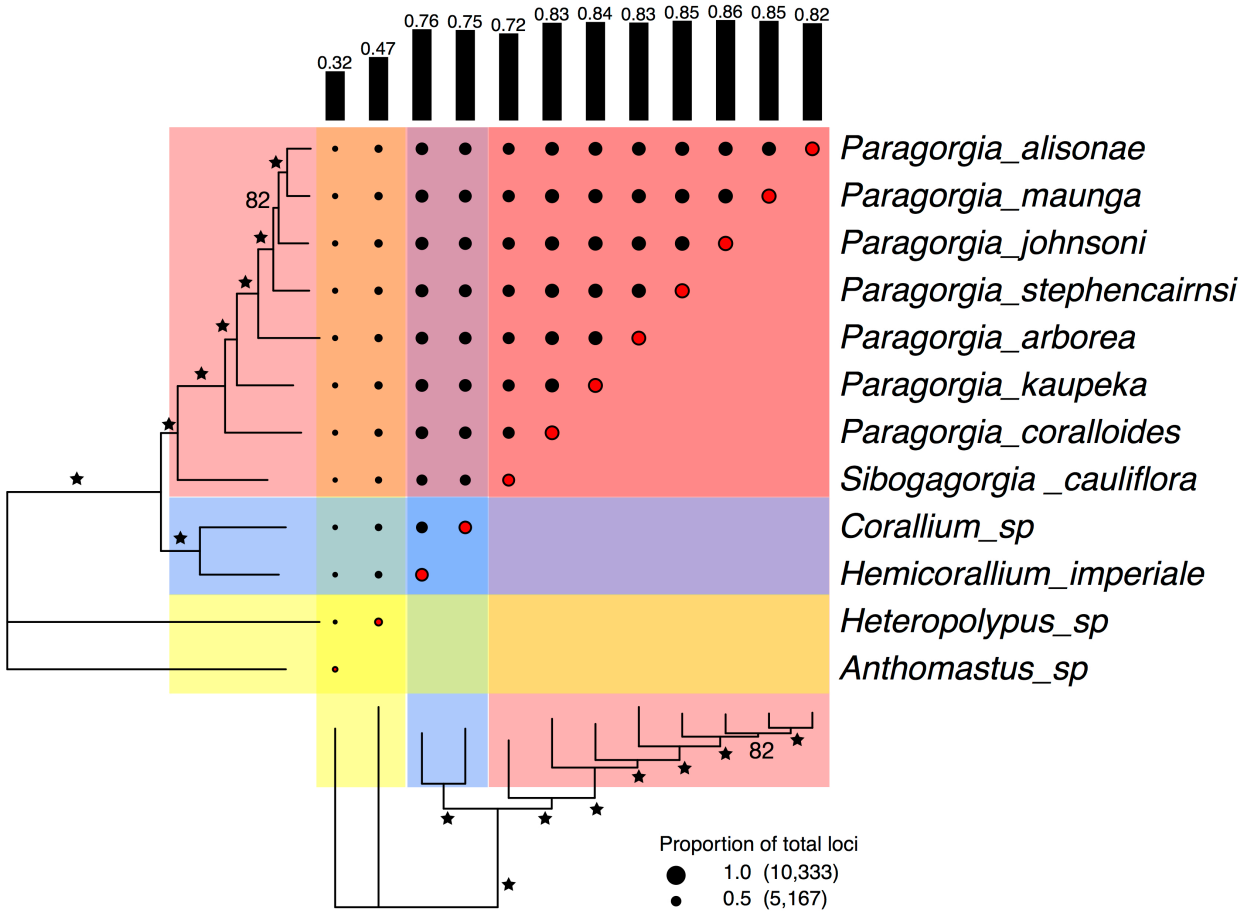


Figure 1. Proportion of loci shared among individuals of the AC clade in the optimal backbone matrix (c 0.80, m 9). Each family is indicated with a different color: red for Paragorgiidae; blue for Coralliidae; and yellow for

Alcyoniidae. Black-filled circles represent the proportion of the total number of loci shared among individuals. Red-filled circles represent the proportion of the total number of loci present in each individual. Circle scale shows the number of loci represented by 1.0 and 0.5 circle sizes. Black vertical bars represent the average proportion of loci shared by each individual. Phylogenetic tree was inferred with RAxML. Stars on the tree represent branch bootstrap support of 100. Smaller bootstrap support values are indicated with numbers. This figure was generated with the package RADami (Hipp *et al.* 2014).

We used the selected optimal loci-clustering parameters to generate the **'PHYLO'** matrix, containing the sequence data of all the 44 octocoral specimens. The use of the parameter value c 0.80 yielded approximately 71 ± 15 thousand loci – with a minimum depth of coverage of 5x and after filtering for paralogs– per specimen (mean depth of clusters used in loci construction was 23 ± 8x) (Table S4). The **PHYLO** matrix contained a total of 5,997 loci that contained data for at least 75% of the specimens (after a second paralog removal). There were 85,293 variable sites in this matrix, of which 53,150 were phylogenetically informative.

RAD-seq data support a fully resolved phylogeny

The phylogenetic analysis of the **PHYLO** concatenated RAD-seq matrix produced a completely resolved evolutionary tree of the AC specimens (Fig. 2). In general, all branches were supported by high (greater than 95) bootstrap values, except for the one supporting the clade of *P. johnsoni, P. alisonae,* and *P. maunga*. Each one of the morphologically identified families, genera, and species in this dataset were monophyletic. The branching pattern of the tree is consistent with an expected transition between coalescent processes among species and genera (long deep branches), and population processes within species (short shallow branches).

The topology of the tree obtained with a traditional '**mitochondrial'** matrix (711 base pairs of the *mtMutS* gene containing 130 variable sites, of which 101 were phylogenetically informative) was incongruent with the **PHYLO** tree (Fig. 2). The **mitochondrial** tree indicated a well-supported alternative divergence order for *P. coralloides* and *P. kaupeka* in the *Paragorgia* clade. In addition, the families Paragorgiidae (bubblegum corals) and Coralliidae (precious corals) were not monophyletic. The bubblegum coral genus *Sibogagorgia* appeared more closely related to the precious corals than to the other bubblegum coral genus *Paragorgia,* and the genera *Corallium* and *Hemicorallium* did not form a clade. However, these alternative relationships were not significantly supported by the bootstrap analysis,. Indeed, a substantial proportion of branches on the **mitochondrial** tree were poorly supported (bootstrap values smaller than 80%).

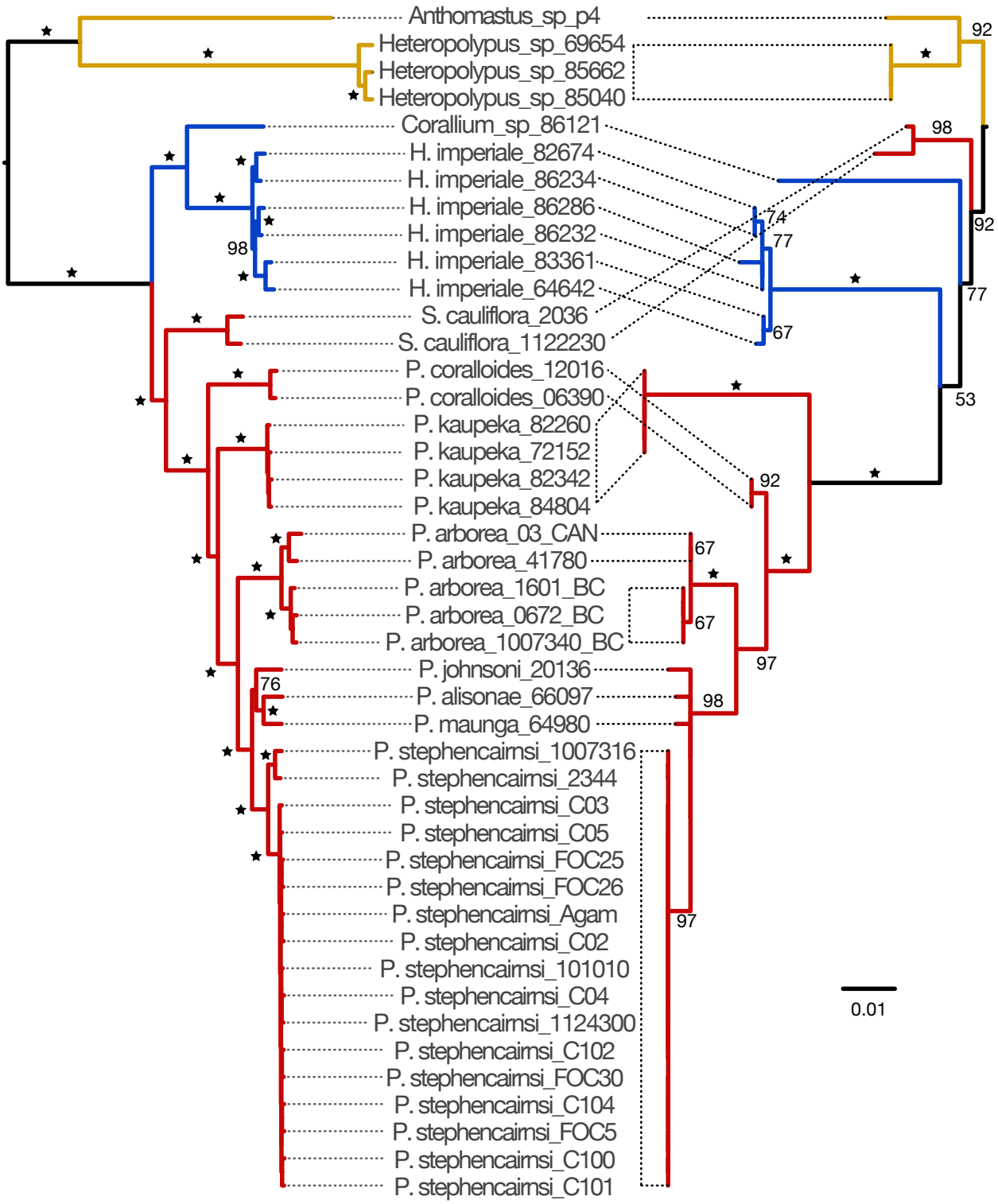


Figure 2. Phylogenetic trees of the AC clade based on RAD-seq and mitochondrial data. Left tree based on the RAD-seq concatenated PHYLO matrix. Right tree based on the mtMutS mitochondrial matrix. Each family is indicated with a different branch color: blue red for Paragorgiidaea; blue for Coralliidae; and yellow for

Alcyoniidae. Phylogenetic trees were inferred with RAxML. Stars on the trees represent bootstrap support of 100. Smaller bootstrap values are indicated in numbers. Scale bar indicates substitutions per site.

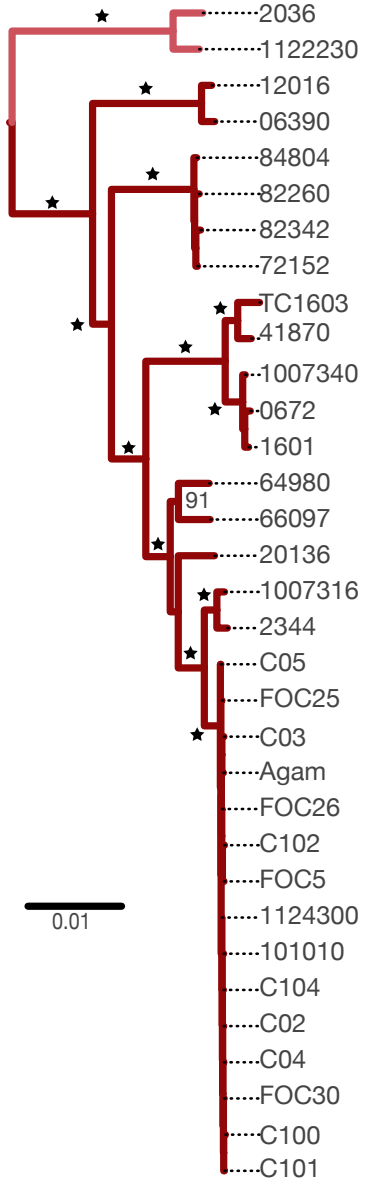
RAD-seq data reveal cryptic genetic diversity

Branch-length differences among individuals, as well as well-supported sub-clades, revealed intraspecific genetic diversity that was undetected by the **mitochondrial** matrix. Two sub-clades were revealed by the phylogenetic analysis of the **PHYLO** matrix in the *P. arborea* and *P. stephencairnsi* clades. The sub-clades in *P. arborea* correspond to a pattern of segregation by geographic location with specimens from the north Pacific in one sub-clade, and specimens from the south Pacific and north Atlantic in the other. Contrastingly, the sub-clades in *P. stephencairnsi* correspond to a pattern of segregation by depth with specimens collected shallower than 350m in one sub-clade, and specimens collected deeper than 1000m in the other.

Current morphological species delimitation is rejected

To evaluate the utility of RAD-seq to perform objective species delimitations in octocorals we focused on specimens the genus *Paragorgia* as it was the best-sampled taxon in our dataset, both in terms of geographic representation and number of morphological species. We used the Bayes Factor Delimitation method with genomic data (BFD*) (Leache *et al.* 2014), which allows for the comparison of conflictive species delimitation models in an explicit multispecies coalescent framework using genome-wide single nucleotide polymorphism (SNP) data. We calculated marginal likelihoods of taxonomy-guided and taxonomy-unguided species delimitation models from a matrix of unlinked SNPs including only specimens of *Paragorgia* (**'PARAGORGIA'** matrix containing 1,203 SNPs present in all individuals). We compared the marginal likelihood estimates of alternative species delimitation models to the null model '**morphid'**, which is based on current morphological species descriptions, using Bayesian factors.

The null model, **morphid,** was rejected in favor of alternative species delimitation models for *Paragorgia* (Fig. 3) (**morphid** was ranked 7th among 10 evaluated models in terms of the marginal likelihood estimate). The '**PABSTE'** model, which proposes 9 species based on the 7 morphological species in the dataset plus splits corresponding to the sub-clades in *P. arborea* and in *P. stephencairnsi,* received decisive support from Bayes factors as the best species delimitation model. The taxonomy-unguided model '**geo'** – which splits the specimens based on the geographic location where they were collected – and the models proposed by the Poisson tree processes (PTP) method based on the **mitochondrial** data matrix, were the lowest ranked and most strongly rejected models overall.



		Species models								
		Taxonomy-informed				Taxonomy-independent				
	morphid	PAB	STE	PAB STE	split PAB	mlPTP mt	bPTP mt	mlPTP rad	bPTP rad	geo
2036	*S. cauliflora*	-	-	-	-	-	-	-	-	-
1122230	*S.cauliflora*	-	-	-	-	-	-	-	-	-
12016	*P. coralloides*	1	1	1	1	8	1	11	8	1
06390	*P. coralloides*	1	1	1	1	7	1	12	9	1
84804	*P. kaupeka*	2	2	2	2	1	17	1	1	2
82260	*P. kaupeka*	2	2	2	2	1	16	1	1	2
82342	*P. kaupeka*	2	2	2	2	1	19	1	1	2
72152	*P. kaupeka*	2	2	2	2	1	20	1	1	2
TC1603	*P. arborea*	3	3	3	3	2	10	7	6	1
41870	*P. arborea*	3	3	3	4	2	10	8	7	2
1007340	*P. arborea*	4	3	4	5	2	13	6	5	3
0672	*P. arborea*	4	3	4	5	2	14	6	5	3
1601	*P. arborea*	4	3	4	5	2	14	6	5	3
64980	*P. maunga*	5	4	5	6	4	3	3	3	2
66097	*P. alisonae*	6	5	6	7	5	4	4	4	2
20136	*P. johnsoni*	7	6	7	8	3	2	2	2	1
1007316	*P. stephencairnsi*	8	7	8	9	6	23	9	12	3
2344	*P. stephencairnsi*	8	7	8	9	6	9	10	13	3
C05	*P. stephencairnsi*	8	8	9	10	6	24	5	10	3
FOC25	*P. stephencairnsi*	8	8	9	10	6	11	5	11	3
C03	*P. stephencairnsi*	8	8	9	10	6	15	5	11	3
Agam	*P. stephencairnsi*	8	8	9	10	6	22	5	11	3
FOC26	*P. stephencairnsi*	8	8	9	10	6	5	5	11	3
C102	*P. stephencairnsi*	8	8	9	10	6	8	5	11	3
FOC5	*P. stephencairnsi*	8	8	9	10	6	24	5	11	3
1124300	*P. stephencairnsi*	8	8	9	10	6	7	5	11	3
101010	*P. stephencairnsi*	8	8	9	10	6	7	5	11	3
C104	*P. stephencairnsi*	8	8	9	10	6	12	5	11	3
C02	*P. stephencairnsi*	8	8	9	10	6	18	5	11	3
C04	*P. stephencairnsi*	8	8	9	10	6	23	5	11	3
FOC30	*P. stephencairnsi*	8	8	9	10	6	21	5	11	3
C100	*P. stephencairnsi*	8	8	9	10	6	6	5	11	3
C101	*P. stephencairnsi*	8	8	9	10	6	11	5	11	3
Number of species		7	8	8	9	10	8	24	12	13	3
Marginal Likelihood		-6240	-6221	-6039	-6014	-6028	-6251	-6243	-6050	-6045	-10296
Rank		7	6	3	1	2	9	8	5	4	10
Bayes Factor		-	-38	-401	-451	-424	23	7	-379	-388	8113

Figure 3. Species delineation hypotheses for *Paragorgia.* Table shows the different species delimitation models for *Paragorgia* evaluated with the BFD* method and their results. *Sibogagorgia* was included as outgroup to root the inferences for *Paragorgia*. Each row indicates a different specimen. Each column indicates a different species

delimitation model. The first column, model morphid, indicates the species identifications based on morphology. For all other models, numbers indicate the species assignments. Bottom rows show the total number of species proposed, the marginal likelihood estimate, and rank for each model. The Bayes factor comparisons were calculated with respect to the null morphid model. Phylogenetic tree on the left, shown only for visual reference, was inferred with the RAD-seq concatenated PARAGORGIID matrix in RAxML. Each genus is indicated with a different branch color: pink for *Sibogagorgia*; and dark red for *Paragorgia*. Stars on the trees represent bootstrap support of 100. Smaller bootstrap values are indicated in numbers. Scale bar indicates substitutions per site.

Concatenated and coalescent species tree analyses are congruent

The topology of the species tree inferred using the SNP **PARAGORGIA** matrix was entirely congruent with the topology generated by the maximum likelihood phylogenetic analysis of the concatenated sequence matrices (Fig. 4). The species tree analysis also greatly improved support for the clade of *P. johnsoni, P. alisonae,* and *P. maunga.* The posterior distribution of species trees indicated a small fraction of conflictive topologies concentrating in this region of the tree.

DISCUSSION

RAD sequencing enables unprecedented phylogenetic resolution

Our analyses of RAD-seq data provide a robust phylogenetic hypothesis for the recalcitrant octocorals in the *Anthomastus-Corallium* clade, a result never achieved before. Moreover, this study, together with the work by Pante *et al.* (2014) in the octocoral genus *Chrysogorgia*, constitute the first applications of RAD-sequencing for phylogenetics and species delimitation in cnidarians. Only a handful of previous studies, using traditional mitochondrial data and the ITS2 and 28S nuclear markers, have attempted to evaluate phylogenetic relationships in the octocoral AC clade (Herrera *et al.* 2010; Ardila *et al.* 2012; Brockman & McFadden 2012; Herrera *et al.* 2012; McFadden & van Ofwegen 2013; Uda *et al.* 2013; Figueroa & Baco 2014). These studies find support for the monophyly of the genus *Paragorgia,* the family Coralliidae, and the sister relationship between the Paragorgiidae and Coralliidae. However, those data do not provide enough phylogenetic resolution to infer the evolutionary relationships among many of the putative morphological species. Furthermore, significant incongruences between mitochondrial and nuclear ITS2 gene trees from AC taxa have been documented (Herrera *et al.* 2010). Here we reproduce similar incongruences when comparing the trees inferred from mitochondrial and RAD-seq datasets (Fig. 2). Likewise, Pante *et al.* (2014) documented marked incongruence between trees inferred from mitochondrial and RAD-seq data in *Chyrsogorgia.* These observations suggest that processes that can

cause gene tree heterogeneity, such as incomplete lineage sorting and horizontal gene transfer (Maddison 1997; Edwards 2009), may be more prevalent in octocorals than previously recognized.

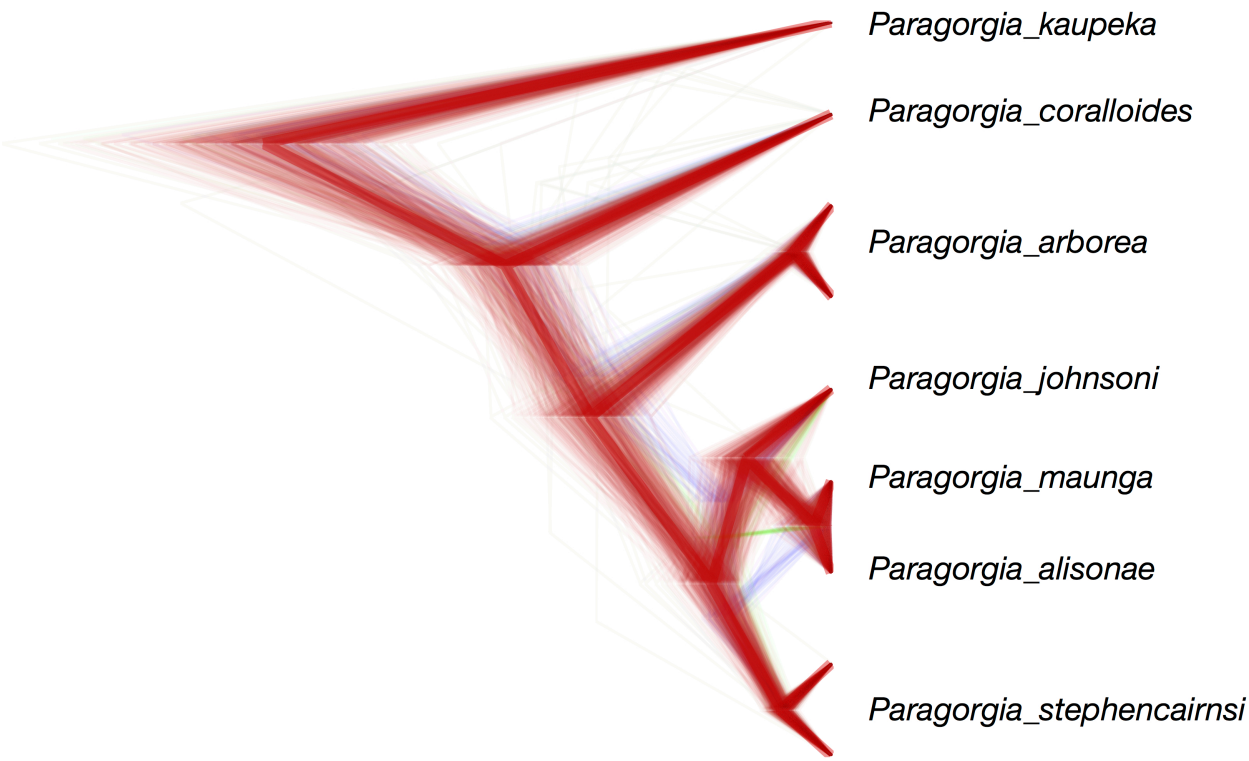


Figure 4. Species tree of *Paragorgia.* This claudogram illustrates the posterior distribution of species trees inferred with SNAPP based on the best species delimitation model PABSTE. High color density is indicative of areas in the species trees with high topology agreement. Different colors represent different topologies. The maximum clade credibility species tree is shown with thicker branches. Trees with the same topology as the maximum clade credibility species tree are colored in red. Trees with different topologies are colored green or blue. With the exception of the branch leading to the clade of *P. johnsoni, P. maunga*, and *P. alisonae,* which has a posterior probability of 0.87, all interior branches have posterior probabilities of 1.0.

All of our analyses based on RAD-seq matrices – varying in taxon coverage, degree of divergence among taxa, proportion of missing data, number of loci, and analysis type (concatenated or species tree) – produced completely congruent trees, which together provide extremely high confidence on the phylogenetic hypothesis inferred for the octocoral AC clade (Figs. 1, 2 and 3). Consequently, we suggest that single marker gene trees in octocorals, particularly from the mitochondria, should not be considered as robust hypotheses of true species phylogenies on their own, without further validation by multiple

informative and independent nuclear loci. We urge systematists to be conservative when making taxonomic rearrangements based on inferences from single-marker data alone.

RAD-seq data is suitable for phylogenetic inference in divergent taxa

Contrary to the currently accepted idea that RAD-seq data are only suitable for taxa with divergence times younger than 60 million years (MY) (Rubin *et al.* 2012), we demonstrate their suitability well beyond this age threshold. Remarkably, we were able to confidently resolve phylogenetic relationships among genera from different families diverging by at least 80 MY in the AC clade. The split between the families Paragorgiidae and Coralliidae has been dated, using coralliid fossils, to be between 80-150 MY old (Ardila *et al.* 2012; Herrera *et al.* 2012). Park *et al.* (2012) estimated the age of the most recent common ancestor of the Coralliidae at approximately 50 MY (25-100 MY 95% confidence region), using independent cnidarian fossils for molecular clock calibration. The split with the genera *Anthomastus* and *Heteropolypus* is likely older than 100 MY. It is without question that, due to evolution at restriction sites, the number of RAD loci among taxa for which orthology can be established decreases rapidly as divergence increases. However, we suggest that the suitability limits of RAD-seq for phylogenetics in divergent taxa cannot be assessed in terms absolute time, but depend on taxon-specific factors such as mutation rate, generation time and effective population size.

Bioinformatic studies addressing the issue of extent of the suitability of RAD-seq for phylogenetic inference have focused mainly on *Drosophila* as study model (Rubin *et al.* 2012; Cariou *et al.* 2013). Longer generation times and lower metabolic rates in taxa like deep-sea corals, relative to those in organisms like *Drosophila,* could cause a reduction in mutation rates (see review by Baer *et al.* (2007)), which may in turn decrease the evolutionary rates at restriction sites and allow for phylogenetic inferences using RAD-seq in situations of deeper divergence. Consistent with this hypothesis, we observe a nucleotide diversity (π) calculated across all octocoral specimens from the **PHYLO** matrix of 0.012 ± 0.002 (considered a minimum since RAD-seq can underestimate diversity (Arnold *et al.* 2013); see Table S5 and Table S6 for individual values), which is significantly lower than the nucleotide diversity in many of the *Drosophila* species included in the bioinformatic studies by Cariou *et al.* (2013) and Rubin *et al.* (2012). Nonetheless, there are other important factors known to influence genetic diversity across species – and likely the evolutionary rate as well. These factors include the effective population size, selection, habitat kind, geographic range, and mating system (Leffler *et al.* 2012). To sum up, we argue that RAD-seq can be successfully used to infer phylogenetic relationships in certain taxa with deeper divergences than previously suggested. This is particularly true when the number of RAD loci is maximized through

the choice of restriction enzymes with higher cutting frequencies in the target taxon (Herrera *et al.* Chapter 2).

RAD-seq allows the formulation of robust species delineations

Our study, the first statistical rigorous test of species hypothesis in octocorals, provides conclusive evidence rejecting the current morphological species delimitation model for the genus *Paragorgia*. We find decisive support for a nested model that combines species boundaries from morphological taxonomy with cryptic diversity linked to environmental variables of geographic location and depth (Figs. 3 and 4). This nested model, proposes 9 species among the examined specimens. Five of these species correspond to the morphological species *P. coralloides, P. kaupeka, P. alisonae, P. johnsoni,* and *P. maunga.* Two splits, corresponding to sub-clades in the morphological species *P. arborea* and in *P. stephencairnsi*, indicate cases of cryptic species.

Herrera *et al.* (2012) found significant genetic differentiation of the north Pacific populations of *P. arborea* relative to the south Pacific, Atlantic and Indian ocean populations, and suggested that these populations may represent sub-species. The north Pacific populations of *P. arborea* were previously defined as a separate species, *P. pacifica,* by Verrill (1922) based on gross colony morphology, but later combined into a single species by Grasshoff (1979). Sánchez (2005) suggested potential small differences in medullar sclerite sizes and ornamentation between north Pacific specimens and specimens form elsewhere. However, we were unable to recognize these morphological differences in the few examined specimens in this study. Nonetheless, based on the decisive support for the split of *P. arborea* from analysis of genome-wide SNP makers indicates, we resurrect the species *Paragorgia pacifica* for the north Pacific populations of formerly *P. arborea*. We find no evidence of cryptic speciation between the north Atlantic and south Pacific *P. arborea* and therefore conclude it should be considered a single species as previously suggested by Herrera *et al.* (2012).

Depth is an important factor contributing to genetic differentiation and formation of species in the ocean, both shallow (Carlon & Budd 2002; Prada & Hellberg 2013) and deep (Miller *et al.* 2011; Jennings *et al.* 2013; Quattrini *et al.* 2013; Glazier & Etter 2014). The observed cryptic differentiation between specimens of *P. stephencairnsi* collected shallower than 350m and deeper than 1000m indicates that depth is also a diversifying force in octocorals from the AC clade, which had gone undetected due to the low variability of traditional sequence data (Herrera *et al.* 2012). The holotype of *P. stephencairnsi* was collected from approximately 350m in the Georgia Strait of British Columbia, overlapping in depth range and geographic region with that of most of the specimens from the shallow sub-clade examined in this

study. Therefore, we propose to conserve that name *P. stephencairnsi* for that shallow sub-clade, and consider the deep sub-clade as a new species.

Other recent species delimitation studies in anthozoan corals have also revealed significant incongruences when comparing morphological and single-locus species delimitation hypotheses (particularly from mitochondrial data) with phylogenetic evidence from multi-locus datasets (Pante *et al.* 2014; Prada *et al.* 2014). In line with the findings of Pante *et al.* (2014), we find that specimens of *Paragorgia* sharing identical *mtMutS* haplotypes can belong to more than one species. Contrastingly, Herrera *et al.* (2012) present strong evidence showing significant mitochondrial haplotype diversity in the south Pacific and north Atlantic populations of *Paragorgia arborea*. Our observations, together with those from the aforementioned studies, constitute compelling evidence indicating that there is no solid basis for the widespread assumption that *mtMutS* haplotypes may be equivalent to individual octocoral species, as proposed by Thoma *et al.* (2009). The analysis with RAD-seq, or alternative genomic multi-locus methods, of a larger number of specimens from diverse geographic locations and depth horizons will likely reveal further cryptic diversity not characterized by mitochondrial haploytypes (see Fig. S2, Fig. S3, and Table S8), and thus further illuminates taxonomy and systematics in this an other groups.

CONCLUSIONS

In this case study we demonstrate the empirical utility of RAD-seq to resolve phylogenetic relationships among divergent and recalcitrant taxa and to objectively infer species boundaries by testing alternative delimitation hypotheses. We were able to make use of RAD-seq to overcome long-standing technical difficulties in octocoral genetics, and to resolve fundamental questions in species definitions and systematics. We show that classic morphological taxonomy can greatly benefit from integrative approaches that provide objective tests to species delimitation hypothesis. Our results pave the way for addressing further questions in biogeography, species ranges, community ecology, population dynamics and evolution of octocorals and other marine taxa. The results from this study also represent a valuable reference resource for the development of tools, such as SNP arrays, that can be used to perform accurate species identifications, and generate species inventories that will aid the design and implementation of conservation and management policies.

METHODS

To perform identifications using current morphological species descriptions we performed colony observations and scanning electron microscopy of sclerites on 44 octocoral specimens from the AC clade (Table S1).

To obtain a genome-wide set of markers that could be useful for phylogenetic inferences of deep-divergent taxa and species delimitation in the AC clade (greater than 100 million years) we performed RAD sequencing with a the 6-cutter restriction enzyme PstI, which is predicted to cut between 32,000 and 110,000 times in the genome of an octocoral (Table S7). This predicted range was obtained using the observed frequency of the PstI recognition sequence, and its probability calculated using a trinucleotide composition model, in the genomes of the cnidarians *Nematostella vectensis, Acropora digitifera, Hydra vulgaris,* and *Alatina moseri* (Herrera *et al.* Chapter 2). Genome size range of 0.3-0.5 pg was used based on observations obtained through flow cytometry in gorgoniid octocorals by Luisa Dueñas at the Universidad de los Andes, Bogotá, Colombia (personal communication). Total genomic DNA was purified from specimens following protocols described in Herrera *et al.* (Chapter 3). Concentration-normalized genomic DNA was submitted to Floragenex Inc (Eugene, OR). for library preparation and RAD sequencing. Libraries were sequenced by 48-multiplex, using 10-base pair barcodes, on a single lane of an Illumina Hi-Seq 2000 sequencer.

To compare the inferences obtained from RAD-seq data with the inferences drawn from traditional genetic barcoding data, we performed targeted sequencing of the mitochondrial *mtMutS* gene — a genetic marker widely used for phylogenetics and species delimitation studies in octocorals. Polymerase chain reactions were carried out following the protocols by Herrera *et al.* (Chapter 3). Primer pairs used for amplifications were AnthoCorMSH (Herrera *et al.* 2010) and Mut-3458R (Sánchez *et al.* 2003). Negative controls were included in every experiment to test for contamination. Purified PCR products were submitted to Eurofins Genomics (Eurofins MWG Operon, Inc.) for sequencing.

RAD-seq data filtering

Sequence reads were de-multiplexed and quality filtered with the process_radtags program from the package Stacks v1.20 (Catchen *et al.* 2013). Barcodes and Illumina adapters were excluded from each read and length was truncated to 91bp (-t 91) Reads with ambiguous bases were discarded (-c). Reads with an average quality score below 10 (-s 10) within a sliding window of 15% of the read length (-w 0.15) were discarded (-r). The rescue barcodes and RAD-tags algorithm was enabled (-r). Additional

filtering, and the clustering within and between individuals to identify loci was performed using the program pyRAD v2.01 (Eaton 2014). Reads with more than 33 bases with a quality score below 20 were discarded.

RAD-seq loci clustering and phylogenetic inference

We investigated different combinations of clustering thresholds (c 0.80, 0.85 and 0.90) and minimum number of taxa per locus (m 4, 6, and 9) in a reduced dataset that included one individual from each of the 12 putative morphological species. The minimum depth of coverage required to build a cluster and the maximum number of shared polymorphic sites in a locus were kept constant at 4 (d) and 3 (p) respectively. Loci sequences were concatenated into combined matrices. We refer to these 9 resulting matrices as the **'backbone'** matrices. Each of the resulting backbone matrices was analyzed in RAxML-HPC2 v8.0 (Stamatakis 2006) for maximum likelihood (ML) phylogenetic tree inference. For this, and all the other phylogenetic analyses in RAxML, we assumed a generalized time-reversible DNA substitution model with a gamma-distributed rate variation across sites (GTR GAMMA). Branch support was assessed by 500 bootstrap replicates.

We selected an optimal combination of loci clustering parameters as the set of parameters that minimized the number of missing data and maximized the number of phylogenetically informative sites while producing a highly supported phylogenetic tree. The optimal set of parameters chosen was a clustering threshold of 80% similarity among sequences (c 0.80) and a minimum coverage of taxa per locus of 75% (m 9). A concatenated matrix containing the sequence data of all the 44 octocoral specimens, denominated **'PHYLO',** was built using this parameter combination (c 0.80, m 33) in pyRAD and subsequently analyzed in RAxML.

Phylogenetic inference with traditional genetic barcoding data

To compare the tree topology obtained from the phylogenetic inferences of the **PHYLO** RAD-seq dataset with traditional genetic barcoding data we analyzed the '**mitochondrial'** dataset (containing the *mtMutS* sequences) using RAxML. These two datasets – **PHYLO** and **mitochondrial** – contain data from the same individuals. To place the specimens from this study in a broader phylogenetic context we also analyzed the **mitochondrial** dataset in RAxML with the addition of *mtMutS* data from 233 additional specimens belonging to the AC clade, as well as outgroups (see Table S8, Fig. S2, and Fig. S3).

Testing species delimitation models for Paragorgia

We constructed 5 taxonomy-guided species delimitation models for *Paragorgia*: **i**) '**morphid'** model: 7 species based on current morphological species descriptions (Sánchez 2005); **ii**) '**PAB'** model: 8 species based on the 7 morphological species plus a split of *P. arborea* based on previous evidence of genetic differentiation of north Pacific populations (Herrera *et al.* 2012); **iii**) '**STE'** model: 8 species based on the 7 morphological species plus a split of *P. stephencairnsi* based on depth differences (specimens collected <350m vs. >1000m), as depth is known to be an important structuring variable in marine taxa (Jennings *et al.* 2013; Prada & Hellberg 2013; Quattrini *et al.* 2013); i**v**) '**PABSTE'** model: 9 species based on the 7 morphological species plus the splits of the **PAB** and **STE** models; **v**) '**splitPAB'** model: 10 species based on the 7 morphological species plus the split of the **STE** model and an additional split in the **PAB** model where *P. arborea* is split in 3 species corresponding to the ocean basin where the specimens were collected (north Pacific, south Pacific and north Atlantic).

We also generated taxonomy-unguided species delimitation models for *Paragorgia* through Bayesian and ML implementations of the Poisson tree processes model (PTP) (available at http://species.h-its.org/ptp/). PTP estimates the number of speciation events in a rooted phylogenetic tree in terms of nucleotide substitutions (Zhang *et al.* 2013). We used PTP to analyze the trees obtained from phylogenetic inferences in RAxML of reduced *mtMutS* and RAD-seq datasets that include only members of the family Paragorgiidae (genera *Paragorgia* and *Sibogagorgia*). The '**PARAGORGIIDAE'** RAD-seq concatenated matrix was generated in pyRAD using a clustering threshold of 80% similarity among sequences (c 0.80) and a minimum coverage of taxa per locus of 100% (m 33). The resulting phylogenetic trees of *Paragorgia* were rooted with the specimens of *Sibogagorgia* and analyzed by the PTP method using a Markov Chain Monte Carlo (MCMC) chain length of 500,000 generations (100 thinning, 25% burnin). We assessed convergence by examining the likelihood trace. The combinations of the ML or Bayesian PTP implementations (mlPTP and bPTP, respectively) with the *mtMutS* or RAD-seq trees of *Paragorgia* resulted in four species delimitation models: **i) 'mlPTP*mt*'** model; **ii) 'bPTP*mt*'** model; **iii) 'mlPTPrad'** model; and **iv) 'bPTPrad'** model. Lastly, because deep-sea corals are known to show genetic differentiation at ocean basin/regional scales (Miller *et al.* 2011; Morrison *et al.* 2011; Herrera *et al.* 2012), we constructed an additional taxonomy-unguided species delimitation model – the '**geo'** model – based on the geographic location where the specimens were collected (north Pacific, south Pacific or north Atlantic ocean basins).

To estimate the marginal likelihood of each species delimitation model we generated a matrix including only specimens of *Paragorgia*, denominated **'PARAGORGIA'** using a clustering threshold of 80% similarity among sequences (c 0.80) and a minimum coverage of taxa per locus of 100% (m 31) in

pyRAD. In contrast to the **backbone**, **PHYLO**, and **PARAGORGIIDAE** RAD-seq matrices, this matrix contained the data of one SNP per locus and not the entire locus sequence. We analyzed these data using the implementation of BFD* in the SNAPP (Bryant *et al.* 2012) plug-in for the program BEAST v2.1.3 (Bouckaert *et al.* 2014). We performed a path-sampling of 48 steps, with a MCMC chain length of 100,000 (10,000 pre-burnin), following the guidelines from Leache *et al.* (2014). Bayesian factors were calculated from the marginal likelihood estimates for each model and compared using the framework proposed by Kass and Raftery (1995)

Species tree inference

To test the tree topology in the genus *Paragorgia* obtained by the phylogenetic analysis of the **PHYLO** and **PARAGORGIIDAE** concatenated matrices we performed a species tree inference from the SNP data in the **PARAGORGIA** matrix using the program SNAPP. This program allows the inference of species trees from unlinked SNP data (only one SNP per locus retained) bypassing the inference of individual gene trees (Bryant *et al.* 2012). We performed 3 independent runs using a MCMC chain length of 10,000,000 (sampling every 1,000 generations; pre-burnin of 1,000) with default prior distributions for coalescence rate, mutation rate and ancestral population size parameters. We assessed convergence to stationary distributions and effective sample sizes >200 after 10% burnin in the program TRACER (Rambaut & Drummond 2007). Species trees in the posterior distribution were summarized with the program DENSITREE v2.01 (Bouckaert 2010).

ACKNOWLEDGEMENTS

This research was supported by the National Geographic Society/Waitt Foundation (W285-13 to SH); the National Oceanic and Atmospheric Administration (NA09OAR4320129 to TS); the National Science Foundation (OCE-1131620 to TS); the National Aeronautics and Space Administration (NNX09AB76G to TS); and the Academic Programs Office (Ocean Ventures Fund to SH), the Ocean Exploration Institute (Fellowship support to TMS) and the Ocean Life Institute of the Woods Hole Oceanographic Institution (WHOI).

Specimens provided by the National Institute of Water and Atmospheric Research (NIWA) were collected under research programs: Kermadec Arc Minerals, funded by the New Zealand Ministry of Business, Innovation & Employment (MBIE), Auckland University, Institute of Geological and Nuclear Science (GNS), and WHOI; Ocean Survey 20/20 funded by Land Information New Zealand; Impact of resource use on vulnerable deep-sea communities (CO1X0906), funded by MBIE; Nascent Inter-Ridge

Volcanic And Neotectonic Activity, funded by the Ministry for Primary Industries (MPI), GNS, MBIE, and the U. of New Hampshire; Scientific Observer Program funded by MPI; and the Joint New Zealand-USA 2005 NOAA Ring of Fire Expedition, part of NIWA's Seamount Program (FRST CO1X0508). For enabling access to key specimens we thank K. Schnabel (NIWA), S. Mills (NIWA), D. Tracey (NIWA), M. Clark (NIWA), A. Rowden (NIWA), S. Cairns (Smithsonian), E. Cordes (Temple U.), A. Quattrini (Temple U.), G. Workman (Department of Fisheries and Oceans Canada - DFO), M. Wyeth (DFO), K. Anderson (DFO), M. Frey (Royal British Columbia Museum - RBCM), H. Gartner (RBCM), L. Watling (U. Hawaii), J. Adkins (CalTech). We thank P. Aldersdale (CSIRO), N. Ardila (ECOMAR) and J. Sanchez (U. Andes) for assistance with morphological identifications. We also thank E. O'Brien (WHOI), D. Forsman (WHOI), J. Fellows (WHOI), J. & S. Schooner, K. Heylar, and N. McDaniel for invaluable assistance during scuba diving fieldwork in British Columbia (DFO scientific license FIN130270). We thank the chief scientists, masters, crew, scientific personnel, and funding agencies of expeditions AT07-35, KOK0506, Lophelia II 2009, RB-0503, TAN1007, TAN1104, TAN1206, and TAN1213. We also thank A. Tarrant and A. Reitzel for providing helpful comments that improved this manuscript.

REFERENCES

Aguilar C, Sánchez JA (2007) Phylogenetic hypotheses of gorgoniid octocorals according to ITS2 and their predicted RNA secondary structures. *Molecular Phylogenetics and Evolution* **43**, 774-786.

Ardila NE, Giribet G, Sanchez JA (2012) A time-calibrated molecular phylogeny of the precious corals: reconciling discrepancies in the taxonomic classification and insights into their evolutionary history. *BMC Evolutionary Biology* **12**.

Arnold B, Corbett-Detig RB, Hartl D, Bomblies K (2013) RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology* **22**, 3179-3190.

Baer CF, Miyamoto MM, Denver DR (2007) Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nature Reviews Genetics* **8**, 619-631.

Baird NA, Etter PD, Atwood TS*, et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**, 3376.

Bayer FM (1956) Octocorallia. In: *Treatise on Invertebrate Paleontology Part F. Coelenterata* (ed. Moore RC), pp. 163-231. Geological Society of America and University of Kansas Press, Lawrence, Kansas.

Berntson EA, Bayer FM, McArthur AG, France SC (2001) Phylogenetic relationships within the Octocorallia (Cnidaria: Anthozoa) based on nuclear 18S rRNA sequences. *Marine Biology* **138**, 235-246.

Bilewitch JP, Degnan SM (2011) A unique horizontal gene transfer event has provided the octocoral mitochondrial genome with an active mismatch repair gene that has potential for an unusual self-contained function. *BMC Evolutionary Biology* **11**, 228.

Bouckaert R, Heled J, Kuhnert D*, et al.* (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* **10**, e1003537.

Bouckaert RR (2010) DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics* **26**, 1372-1373.

Brockman SA, McFadden CS (2012) The mitochondrial genome of *Paraminabea aldersladei* (Cnidaria: Anthozoa: Octocorallia) supports intramolecular recombination as the primary mechanism of gene rearrangement in octocoral mitochondrial genomes. *Genome Biology and Evolution* **4**, 994-1006.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A (2012) Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution* **29**, 1917-1932.

Cairns S (2007) Deep-water corals: An overview with special reference to diversity and distribution of deep-water scleractinian corals. *Bulletin of Marine Science* **81**, 311-322.

Cariou M, Duret L, Charlat S (2013) Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. *Ecology and Evolution* **3**, 846-852.

Carlon DB, Budd AF (2002) Incipient speciation across a depth gradient in a scleractinian coral? *Evolution* **56**, 2227-2242.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**, 3124-3140.

CITES (2014) Appendices I, II and III. Convention on International Trade in Endangered Species of wild fauna and flora, , http://www.cites.org/sites/default/files/eng/app/2014/E-Appendices-2014-09-14.pdf.

Clark AG, Eisen MB, Smith DR*, et al.* (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203-218.

Concepcion GT, Crepeau MW, Wagner D, Kahng SE, Toonen RJ (2007) An alternative to ITS, a hypervariable, single-copy nuclear intron in corals, and its use in detecting cryptic species within the octocoral genus *Carijoa*. *Coral reefs* **27**, 323-336.

Cruaud A, Gautier M, Galan M*, et al.* (2014) Empirical assessment of RAD sequencing for interspecific phylogeny. *Molecular Biology and Evolution* **31**, 1272-1274.

Daly M, Brugler MR, Cartwright P*, et al.* (2007) The phylum Cnidaria: A review of phylogenetic patterns and diversity 300 years after Linnaeus. *Zootaxa* **1668**, 127-182.

De Queiroz K (2007) Species concepts and species delimitation. *Systematic Biology* **56**, 879-886.

Dobzhansky T (1970) *Genetics of the Evolutionary Process* Columbia University Press, New York,.

Dueñas LF, Alderslade P, Sánchez JA (2014) Molecular systematics of the deep-sea bamboo corals (Octocorallia: Isididae: Keratoisidinae) from New Zealand with descriptions of two new species of *Keratoisis*. *Molecular Phylogenetics and Evolution* **74**, 15-28.

Dueñas LF, Sánchez JA (2009) Character lability in deep-sea bamboo corals (Octocorallia, Isididae, Keratoisidinae). *Marine Ecology Progress Series* **397**, 11-23.

Eaton DA (2014) PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* **30**, 1844-1849.

Eaton DAR, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: An example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology* **62**, 689-706.

Edwards SV (2009) Is a new and general theory of molecular systematics emerging? *Evolution* **63**, 1-19.

Emerson KJ, Merz CR, Catchen JM*, et al.* (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 16196-16200.

Escudero M, Eaton DAR, Hahn M, Hipp AL (2014) Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: A case study in *Carex* (Cyperaceae). *Molecular Phylogenetics and Evolution* **79**, 359-367.

Figueroa DF, Baco AR (2014) Complete mitochondrial genomes elucidate phylogenetic relationships of the deep-sea octocoral families Coralliidae and Paragorgiidae. *Deep-Sea Research Part Ii-Topical Studies in Oceanography* **99**, 83-91.

France SC (2007) Genetic analysis of bamboo corals (Cnidaria : Octocorallia : Isididae): Does lack of colony branching distinguish *Lepidisis* from *Keratoisis*? In: *Bulletin of Marine Science*, pp. 323-333.

France SC, Hoover LL, Hoover LL (2002) DNA sequences of the mitochondrial COI gene have low levels of divergence among deep-sea octocorals (Cnidaria: Anthozoa). *Hydrobiologia* **471**, 149-155.

Fujisawa T, Barraclough TG (2013) Delimiting species using single-locus data and the generalized mixed Yule coalescent approach: A revised method and evaluation on simulated data sets. *Systematic Biology* **62**, 707-724.

Glazier AE, Etter RJ (2014) Cryptic speciation along a bathymetric gradient. *Biological Journal of the Linnean Society* **113**, 897–913.

Grajales A, Aguilar C, Sánchez JA (2007) Phylogenetic reconstruction using secondary structures of Internal Transcribed Spacer 2 (ITS2, rDNA): finding the molecular and morphological gap in Caribbean gorgonian corals. *BMC Evolutionary Biology* **7**, 90.

Grasshoff M (1979) Zur bipolaren verbreitung der oktokoralle *Paragorgia arborea* (Cnidaria: Anthozoa: Scleraxonia). *Senckenbergiana Maritima* **11**, 115-137.

Grummer JA, Bryson RW, Reeder TW (2014) Species delimitation using Bayes factors: Simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). *Systematic Biology* **63**, 119-133.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B-Biological Sciences* **270**, 313-321.

Hellberg ME (2006) No variation and low synonymous substitution rates in coral mtDNA despite high nuclear variation. *BMC Evolutionary Biology* **6**.

Herrera S, Baco A, Sánchez JA (2010) Molecular systematics of the bubblegum coral genera (Paragorgiidae, Octocorallia) and description of a new deep-sea species. *Molecular Phylogenetics and Evolution* **55**, 123-135.

Herrera S, Reyes-Herrera PH, Shank TM (Chapter 2) Genome-wide predictability of restriction sites across the eukaryotic tree of life.

Herrera S, Shank TM, Sánchez JA (2012) Spatial and temporal patterns of genetic variation in the widespread antitropical deep-sea coral *Paragorgia arborea*. *Molecular Ecology* **21**, 6053-6067.

Herrera S, Watanabe H, Shank T (Chapter 3) Evolutionary and biogeographical patterns of barnacles from deep-sea hydrothermal vents.

Hipp AL, Eaton DAR, Cavender-Bares J*, et al.* (2014) A framework phylogeny of the american oak clade based on sequenced RAD data. *PLoS One* **9**, e93975.

ICES (2013) Assessment of the list of VME indicator species and elements. International Council for the Exploration of the Sea, http://www.ices.dk/sites/pub/Publication Reports/Advice/2013/Special requests/NEAFC_VME_ indicator_ species_ and_elements.pdf.

Jennings RM, Etter RJ, Ficarra L (2013) Population differentiation and species formation in the deep sea: the potential role of environmental gradients and depth. *PLoS One* **8**, e77594.

Jones JC, Fan S, Franchini P, Schartl M, Meyer A (2013) The evolutionary history of *Xiphophorus* fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. *Molecular Ecology* **22**, 2986-3001.

Kass RE, Raftery AE (1995) Bayes Factors. *Journal of the American Statistical Association* **90**, 773-795.

Leache AD, Fujita MK, Minin VN, Bouckaert RR (2014) Species Delimitation using Genome-Wide SNP Data. *Systematic Biology* **63**, 534-542.

Leffler EM, Bullaughey K, Matute DR*, et al.* (2012) Revisiting an old riddle: What determines genetic diversity levels within species? *PLoS Biology* **10**, e1001388.

Maddison WP (1997) Gene trees in species trees. *Systematic Biology* **46**, 523-536.

Mayr E (1942) *Systematics and the Origin of Species from the Viewpoint of a Zoologist* Columbia University Press, New York,.

Mcfadden CS, Benayahu Y, Pante E*, et al.* (2010a) Limitations of mitochondrial gene barcoding in Octocorallia. *Molecular Ecology Resources* **11**, 19-31.

McFadden CS, France SC, Sánchez JA, Alderslade P (2006) A molecular phylogenetic analysis of the Octocorallia (Cnidaria: Anthozoa) based on mitochondrial protein-coding sequences. *Molecular Phylogenetics and Evolution* **41**, 513-527.

McFadden CS, Sanchez JA, France SC (2010b) Molecular Phylogenetic Insights into the Evolution of Octocorallia: A Review. *Integrative and Comparative Biology* **50**, 389-410.

Mcfadden CS, Tullis ID, Breton Hutchinson M, Winner K, Sohm JA (2004) Variation in Coding (NADH Dehydrogenase Subunits 2, 3, and 6) and Noncoding Intergenic Spacer Regions of the Mitochondrial Genome in Octocorallia (Cnidaria: Anthozoa). *Marine Biotechnology* **6**, 516-526.

McFadden CS, van Ofwegen LP (2013) Molecular phylogenetic evidence supports a new family of octocorals and a new genus of Alcyoniidae (Octocorallia, Alcyonacea). *Zookeys* **346**, 59-83.

Miller KJ, Rowden AA, Williams A, Haussermann V (2011) Out of their depth? Isolated deep populations of the cosmopolitan coral *Desmophyllum dianthus* may be highly vulnerable to environmental change. *PLoS One* **6**, e19004.

Morrison CL, Ross SW, Nizinski MS*, et al.* (2011) Genetic discontinuity among regional populations of *Lophelia pertusa* in the North Atlantic Ocean. *Conservation Genetics* **12**, 713-729.

Nadeau NJ, Martin SH, Kozak KM*, et al.* (2012) Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Molecular Ecology* **22**, 814-826.

Pante E, Abdelkrim J, Viricel A, Gey D, France S (2014) Use of RAD sequencing for delimiting species. *Heredity*.

Park E, Hwang DS, Lee JS*, et al.* (2012) Estimation of divergence times in cnidarian evolution based on mitochondrial protein-coding genes and the fossil record. *Molecular Phylogenetics and Evolution* **62**, 329-345.

Pons J, Barraclough TG, Gomez-Zurita J*, et al.* (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology* **55**, 595-609.

Prada C, DeBiasse MB, Neigel JE*, et al.* (2014) Genetic species delineation among branching Caribbean *Porites* corals. *Coral Reefs* **33**, 1019-1030.

Prada C, Hellberg ME (2013) Long prereproductive selection and divergence by depth in a Caribbean candelabrum coral. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 3961-3966.

Quattrini AM, Georgian SE, Byrnes L*, et al.* (2013) Niche divergence by deep-sea octocorals in the genus *Callogorgia* across the continental slope of the Gulf of Mexico. *Molecular Ecology* **22**, 4123-4140.

Rambaut A, Drummond AJ (2007) Tracer v1.4, Available from http://beast.bio.ed.ac.uk/Tracer.

Reitzel AM, Herrera S, Layden MJ, Martindale MQ, Shank TM (2013) Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular Ecology* **22**, 2953-2970.

Roberts JM, Cairns SD (2014) Cold-water corals in a changing ocean. *Current Opinion in Environmental Sustainability* **7**, 118-126.

Roberts JM, Wheeler A, Freiwald AR, Cairns SD (2009) *Cold-Water Corals : The Biology and Geology of Deep-Sea Coral Habitats* Cambridge University Press, Cambridge, UK ; New York.

Rubin BE, Ree RH, Moreau CS (2012) Inferring phylogenies from RAD sequence data. *PLoS One* **7**, e33394.

Sánchez JA (2005) Systematics of the bubblegum corals (Cnidaria: Octocorallia: Paragorgiidae) with description of new species from New Zealand and the Eastern Pacific. *Zootaxa* **1014**, 1-72.

Sanchez JA, Dorado D (2008) Intragenomic ITS2 variation in Caribbean seafans. In: *Proceedings of the 11th International Coral Reef Symposium*, pp. 1383-1387, Ft. Lauderdale.

Sánchez JA, Lasker HR, Taylor DJ (2003) Phylogenetic analyses among octocorals (Cnidaria): mitochondrial and nuclear DNA sequences (lsu-rRNA, 16S and ssu-rRNA, 18S) support two convergent clades of branching gorgonians. *Molecular Phylogenetics and Evolution* **29**, 31-42.

Seetharam AS, Stuart GW (2013) Whole genome phylogeny for 21 *Drosophila* species using predicted 2b-RAD fragments. *PeerJ* **1**, e226.

Shearer TL, Van Oppen MJH, Romano SL, Worheide G (2002) Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). *Molecular Ecology* **11**, 2475-2487.

Smith P, McVeagh S, Mingoia J, France S (2004) Mitochondrial DNA sequence variation in deep-sea bamboo coral (Keratoisidinae) species in the southwest and northwest Pacific Ocean. *Marine Biology* **144**, 253-261.

Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690.

Thoma J, Pante E, Brugler M, France S (2009) Deep-sea octocorals and antipatharians show no evidence of seamount-scale endemism in the NW Atlantic. *Marine Ecology Progress Series* **397**, 25-35.

Uda K, Komeda Y, Fujita T*, et al.* (2013) Complete mitochondrial genomes of the Japanese pink coral (*Corallium elatius*) and the Mediterranean red coral (*Corallium rubrum*): a reevaluation of the phylogeny of the family Coralliidae based on molecular data. *Comparative Biochemistry and Physiology D-Genomics & Proteomics* **8**, 209-219.

Verrill AE (1922) Part G: Alcyonaria and Actiniaria. *Report of the Canadian Arctic Expedition 1913-18* **8**, 1-164.

Wagner CE, Keller I, Wittwer S*, et al.* (2012) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology* **22**, 787-798.

Wating L, France SC, Pante E, Simpson A (2011) Biology of deep-water octocorals. *Advances in Marine Biology, Vol 60* **60**, 41-122.

Yang ZH, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 9264-9269.

Zapata F, Jimenez I (2012) Species delimitation: Inferring gaps in morphology across geography. *Systematic Biology* **61**, 179-194.

Zhang J, Kapli P, Pavlidis P, Stamatakis A (2013) A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* **29**, 2869-2876.

SUPPLEMENTARY MATERIALS

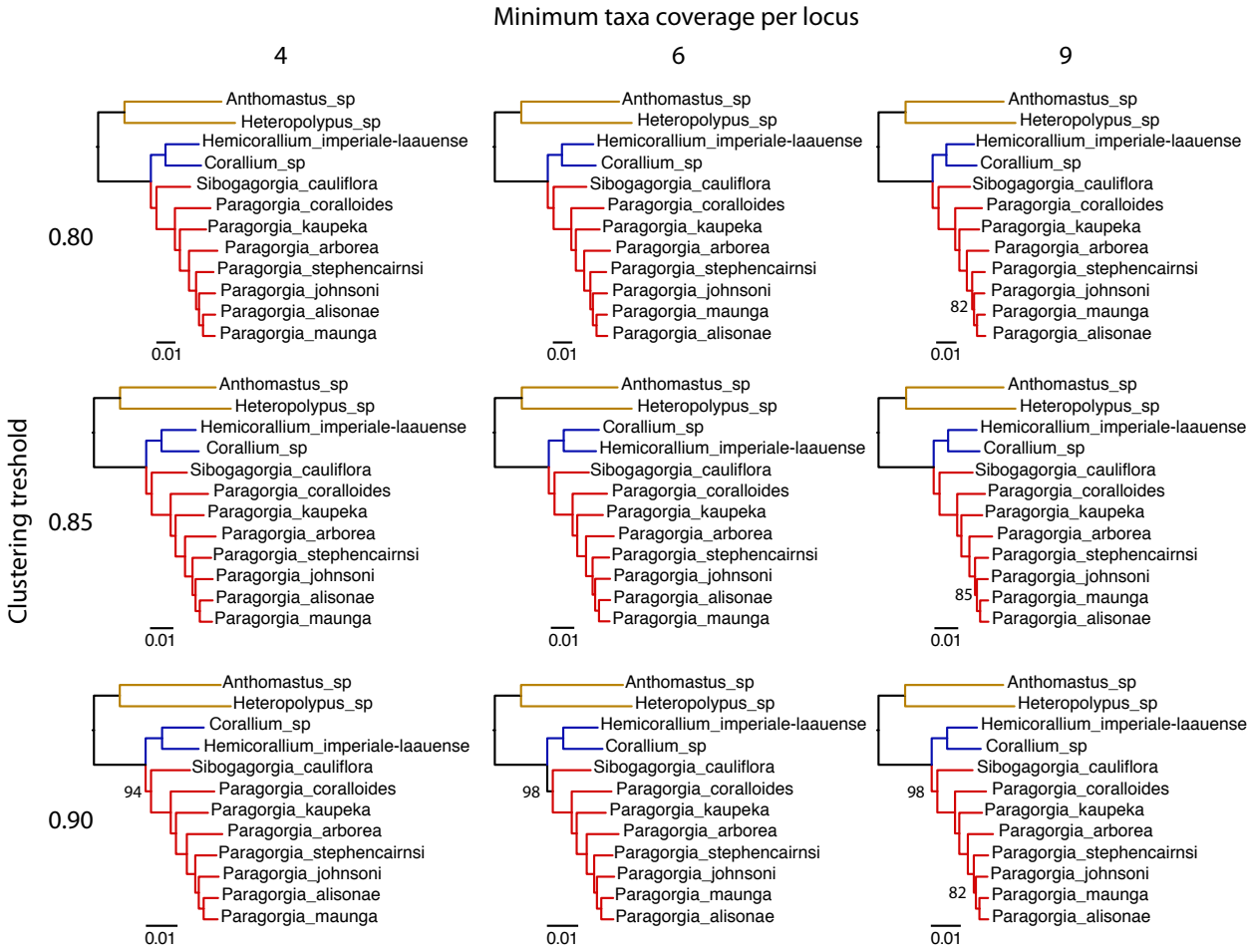
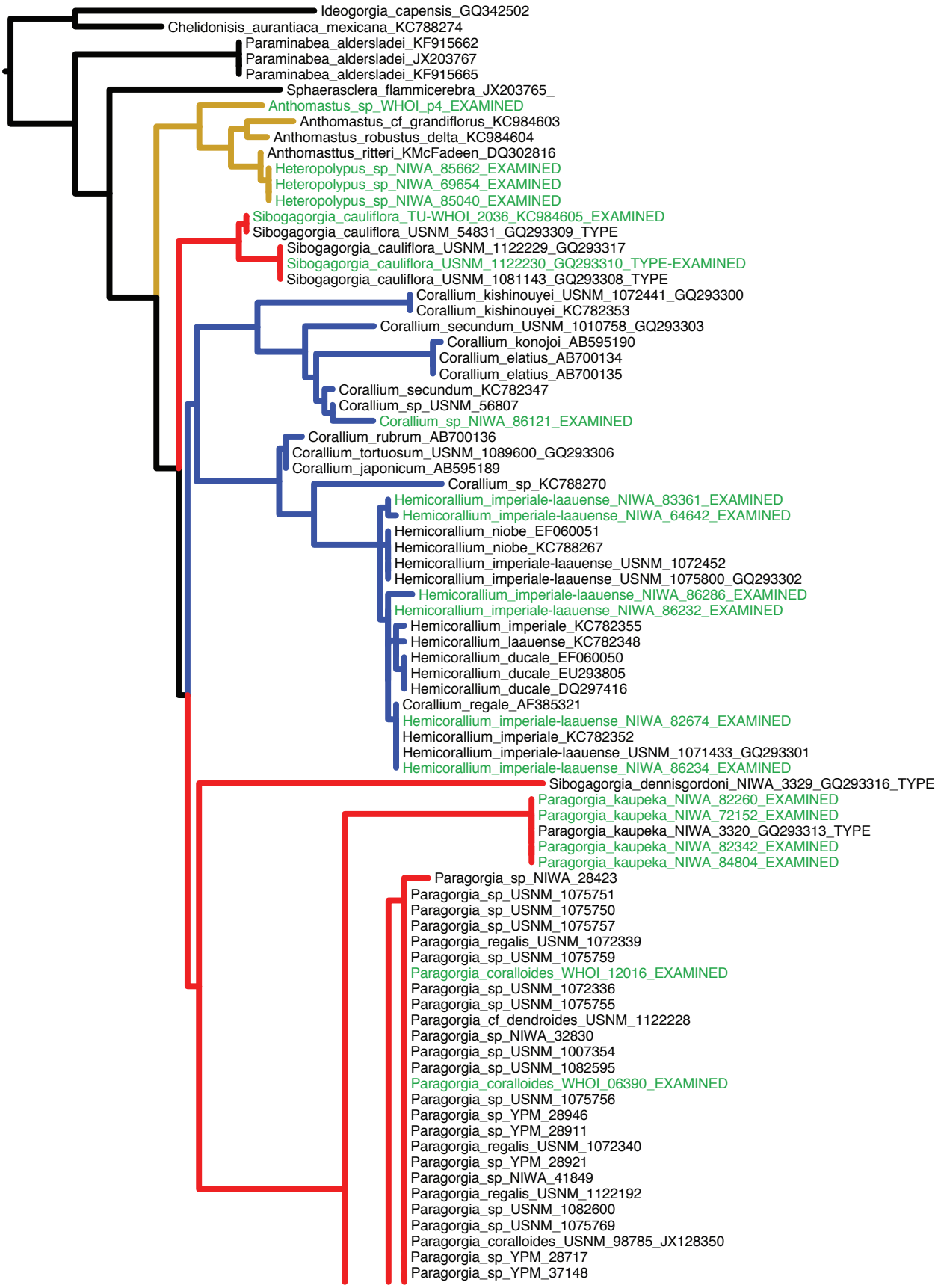
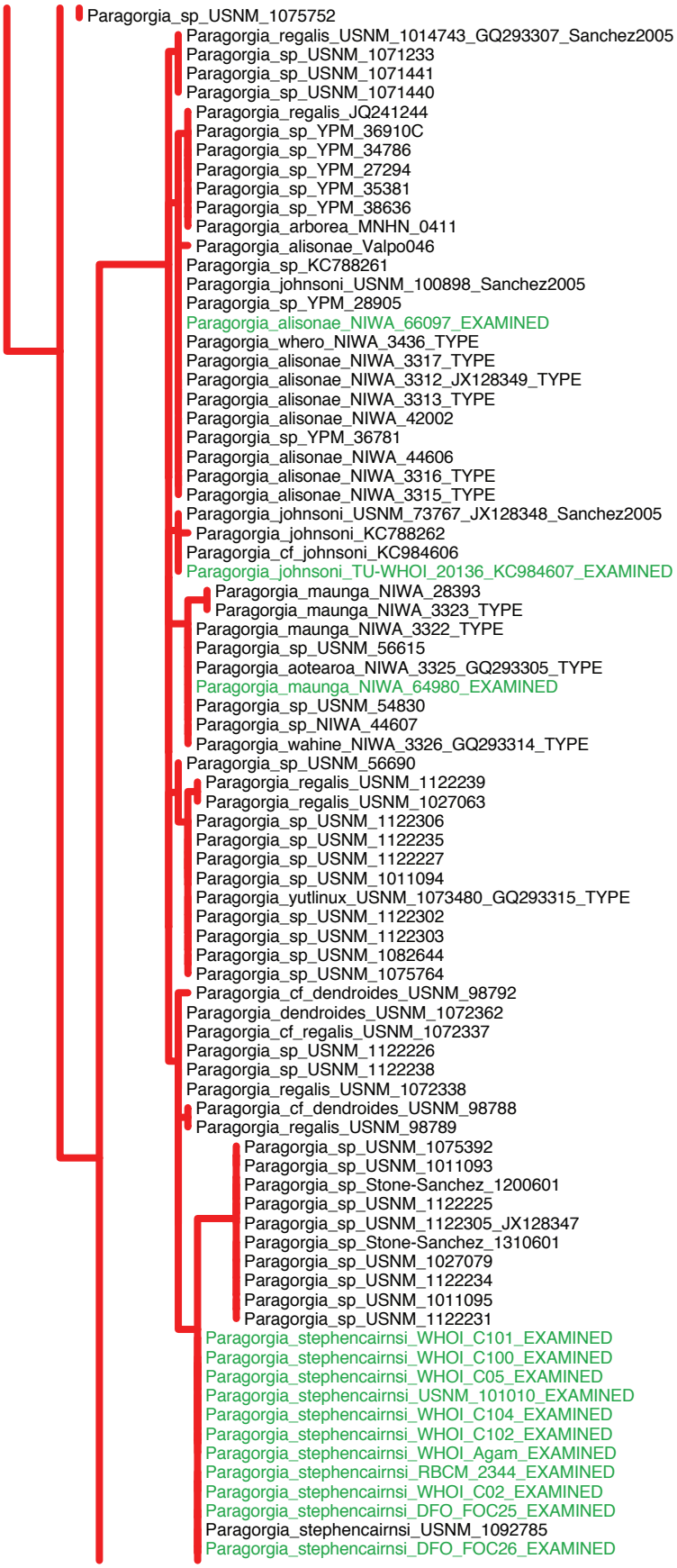
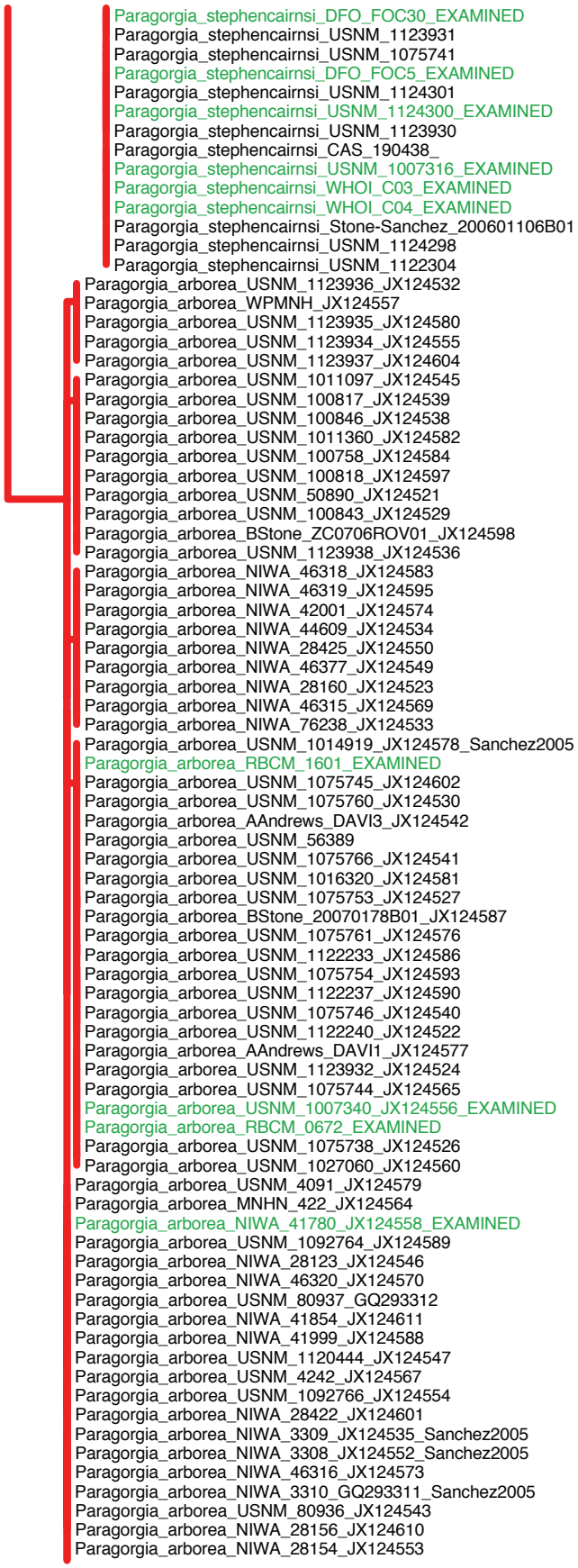


Figure S1. Phylogenetic trees based on backbone matrices. Trees inferred from the 9 backbone RAD-seq matrices built with different parameters of clustering threshold (c 0.80, 0.85 and 0.90; indicated by vertical labels) and minimum number of taxa per locus (m 4, 6, and 9; indicated by horizontal labels). Each family is indicated with a different branch color: red for Paragorgiidae; blue for Coralliidae; and yellow for Alcyoniidae. Trees were inferred with RAxML. All interior branches have bootstrap support values of 100, except for those shown. Scale bars indicate substitutions per site.


Ideogorgia_capensis_GQ342502
Chelidonisis_aurantiaca_mexicana_KC788274
Paraminabea_aldersladei_KF915662
Paraminabea_aldersladei_JX203767
Paraminabea_aldersladei_KF915665
Sphaerasclera_flammicerebra_JX203765_
Anthomastus_sp_WHOI_p4_EXAMINED
Anthomastus_cf_grandiflorus_KC984603
Anthomastus_robustus_delta_KC984604
Anthomasttus_ritteri_KMcFadeen_DQ302816
Heteropolypus_sp_NIWA_85662_EXAMINED
Heteropolypus_sp_NIWA_69654_EXAMINED
Heteropolypus_sp_NIWA_85040_EXAMINED
Sibogagorgia_cauliflora_TU-WHOI_2036_KC984605_EXAMINED
Sibogagorgia_cauliflora_USNM_54831_GQ293309_TYPE
Sibogagorgia_cauliflora_USNM_1122229_GQ293317
Sibogagorgia_cauliflora_USNM_1122230_GQ293310_TYPE-EXAMINED
Sibogagorgia_cauliflora_USNM_1081143_GQ293308_TYPE
Corallium_kishinouyei_USNM_1072441_GQ293300
Corallium_kishinouyei_KC782353
Corallium_secundum_USNM_1010758_GQ293303
Corallium_konojoi_AB595190
Corallium_elatius_AB700134
Corallium_elatius_AB700135
Corallium_secundum_KC782347
Corallium_sp_USNM_56807
Corallium_sp_NIWA_86121_EXAMINED
Corallium_rubrum_AB700136
Corallium_tortuosum_USNM_1089600_GQ293306
Corallium_japonicum_AB595189
Corallium_sp_KC788270
Hemicorallium_imperiale-laauense_NIWA_83361_EXAMINED
Hemicorallium_imperiale-laauense_NIWA_64642_EXAMINED
Hemicorallium_niobe_EF060051
Hemicorallium_niobe_KC788267
Hemicorallium_imperiale-laauense_USNM_1072452
Hemicorallium_imperiale-laauense_USNM_1075800_GQ293302
Hemicorallium_imperiale-laauense_NIWA_86286_EXAMINED
Hemicorallium_imperiale-laauense_NIWA_86232_EXAMINED
Hemicorallium_imperiale_KC782355
Hemicorallium_laauense_KC782348
Hemicorallium_ducale_EF060050
Hemicorallium_ducale_EU293805
Hemicorallium_ducale_DQ297416
Corallium_regale_AF385321
Hemicorallium_imperiale-laauense_NIWA_82674_EXAMINED
Hemicorallium_imperiale_KC782352
Hemicorallium_imperiale-laauense_USNM_1071433_GQ293301
Hemicorallium_imperiale-laauense_NIWA_86234_EXAMINED
Sibogagorgia_dennisgordoni_NIWA_3329_GQ293316_TYPE
Paragorgia_kaupeka_NIWA_82260_EXAMINED
Paragorgia_kaupeka_NIWA_72152_EXAMINED
Paragorgia_kaupeka_NIWA_3320_GQ293313_TYPE
Paragorgia_kaupeka_NIWA_82342_EXAMINED
Paragorgia_kaupeka_NIWA_84804_EXAMINED
Paragorgia_sp_NIWA_28423
Paragorgia_sp_USNM_1075751
Paragorgia_sp_USNM_1075750
Paragorgia_sp_USNM_1075757
Paragorgia_regalis_USNM_1072339
Paragorgia_sp_USNM_1075759
Paragorgia_coralloides_WHOI_12016_EXAMINED
Paragorgia_sp_USNM_1072336
Paragorgia_sp_USNM_1075755
Paragorgia_cf_dendroides_USNM_1122228
Paragorgia_sp_NIWA_32830
Paragorgia_sp_USNM_1007354
Paragorgia_sp_USNM_1082595
Paragorgia_coralloides_WHOI_06390_EXAMINED
Paragorgia_sp_USNM_1075756
Paragorgia_sp_YPM_28946
Paragorgia_sp_YPM_28911
Paragorgia_regalis_USNM_1072340
Paragorgia_sp_YPM_28921
Paragorgia_sp_NIWA_41849
Paragorgia_regalis_USNM_1122192
Paragorgia_sp_USNM_1082600
Paragorgia_sp_USNM_1075769
Paragorgia_coralloides_USNM_98785_JX128350
Paragorgia_sp_YPM_28717
Paragorgia_sp_YPM_37148


Paragorgia_sp_USNM_1075752
Paragorgia_regalis_USNM_1014743_GQ293307_Sanchez2005
Paragorgia_sp_USNM_1071233
Paragorgia_sp_USNM_1071441
Paragorgia_sp_USNM_1071440
Paragorgia_regalis_JQ241244
Paragorgia_sp_YPM_36910C
Paragorgia_sp_YPM_34786
Paragorgia_sp_YPM_27294
Paragorgia_sp_YPM_35381
Paragorgia_sp_YPM_38636
Paragorgia_arborea_MNHN_0411
Paragorgia_alisonae_Valpo046
Paragorgia_sp_KC788261
Paragorgia_johnsoni_USNM_100898_Sanchez2005
Paragorgia_sp_YPM_28905
Paragorgia_alisonae_NIWA_66097_EXAMINED
Paragorgia_whero_NIWA_3436_TYPE
Paragorgia_alisonae_NIWA_3317_TYPE
Paragorgia_alisonae_NIWA_3312_JX128349_TYPE
Paragorgia_alisonae_NIWA_3313_TYPE
Paragorgia_alisonae_NIWA_42002
Paragorgia_sp_YPM_36781
Paragorgia_alisonae_NIWA_44606
Paragorgia_alisonae_NIWA_3316_TYPE
Paragorgia_alisonae_NIWA_3315_TYPE
Paragorgia_johnsoni_USNM_73767_JX128348_Sanchez2005
Paragorgia_johnsoni_KC788262
Paragorgia_cf_johnsoni_KC984606
Paragorgia_johnsoni_TU-WHOI_20136_KC984607_EXAMINED
Paragorgia_maunga_NIWA_28393
Paragorgia_maunga_NIWA_3323_TYPE
Paragorgia_maunga_NIWA_3322_TYPE
Paragorgia_sp_USNM_56615
Paragorgia_aotearoa_NIWA_3325_GQ293305_TYPE
Paragorgia_maunga_NIWA_64980_EXAMINED
Paragorgia_sp_USNM_54830
Paragorgia_sp_NIWA_44607
Paragorgia_wahine_NIWA_3326_GQ293314_TYPE
Paragorgia_sp_USNM_56690
Paragorgia_regalis_USNM_1122239
Paragorgia_regalis_USNM_1027063
Paragorgia_sp_USNM_1122306
Paragorgia_sp_USNM_1122235
Paragorgia_sp_USNM_1122227
Paragorgia_sp_USNM_1011094
Paragorgia_yutlinux_USNM_1073480_GQ293315_TYPE
Paragorgia_sp_USNM_1122302
Paragorgia_sp_USNM_1122303
Paragorgia_sp_USNM_1082644
Paragorgia_sp_USNM_1075764
Paragorgia_cf_dendroides_USNM_98792
Paragorgia_dendroides_USNM_1072362
Paragorgia_cf_regalis_USNM_1072337
Paragorgia_sp_USNM_1122226
Paragorgia_sp_USNM_1122238
Paragorgia_regalis_USNM_1072338
Paragorgia_cf_dendroides_USNM_98788
Paragorgia_regalis_USNM_98789
Paragorgia_sp_USNM_1075392
Paragorgia_sp_USNM_1011093
Paragorgia_sp_Stone-Sanchez_1200601
Paragorgia_sp_USNM_1122225
Paragorgia_sp_USNM_1122305_JX128347
Paragorgia_sp_Stone-Sanchez_1310601
Paragorgia_sp_USNM_1027079
Paragorgia_sp_USNM_1122234
Paragorgia_sp_USNM_1011095
Paragorgia_sp_USNM_1122231
Paragorgia_stephencairnsi_WHOI_C101_EXAMINED
Paragorgia_stephencairnsi_WHOI_C100_EXAMINED
Paragorgia_stephencairnsi_WHOI_C05_EXAMINED
Paragorgia_stephencairnsi_USNM_101010_EXAMINED
Paragorgia_stephencairnsi_WHOI_C104_EXAMINED
Paragorgia_stephencairnsi_WHOI_C102_EXAMINED
Paragorgia_stephencairnsi_WHOI_Agam_EXAMINED
Paragorgia_stephencairnsi_RBCM_2344_EXAMINED
Paragorgia_stephencairnsi_WHOI_C02_EXAMINED
Paragorgia_stephencairnsi_DFO_FOC25_EXAMINED
Paragorgia_stephencairnsi_USNM_1092785
Paragorgia_stephencairnsi_DFO_FOC26_EXAMINED


Paragorgia_stephencairnsi_DFO_FOC30_EXAMINED
Paragorgia_stephencairnsi_USNM_1123931
Paragorgia_stephencairnsi_USNM_1075741
Paragorgia_stephencairnsi_DFO_FOC5_EXAMINED
Paragorgia_stephencairnsi_USNM_1124301
Paragorgia_stephencairnsi_USNM_1124300_EXAMINED
Paragorgia_stephencairnsi_USNM_1123930
Paragorgia_stephencairnsi_CAS_190438_
Paragorgia_stephencairnsi_USNM_1007316_EXAMINED
Paragorgia_stephencairnsi_WHOI_C03_EXAMINED
Paragorgia_stephencairnsi_WHOI_C04_EXAMINED
Paragorgia_stephencairnsi_Stone-Sanchez_200601106B01
Paragorgia_stephencairnsi_USNM_1124298
Paragorgia_stephencairnsi_USNM_1122304
Paragorgia_arborea_USNM_1123936_JX124532
Paragorgia_arborea_WPMNH_JX124557
Paragorgia_arborea_USNM_1123935_JX124580
Paragorgia_arborea_USNM_1123934_JX124555
Paragorgia_arborea_USNM_1123937_JX124604
Paragorgia_arborea_USNM_1011097_JX124545
Paragorgia_arborea_USNM_100817_JX124539
Paragorgia_arborea_USNM_100846_JX124538
Paragorgia_arborea_USNM_1011360_JX124582
Paragorgia_arborea_USNM_100758_JX124584
Paragorgia_arborea_USNM_100818_JX124597
Paragorgia_arborea_USNM_50890_JX124521
Paragorgia_arborea_USNM_100843_JX124529
Paragorgia_arborea_BStone_ZC0706ROV01_JX124598
Paragorgia_arborea_USNM_1123938_JX124536
Paragorgia_arborea_NIWA_46318_JX124583
Paragorgia_arborea_NIWA_46319_JX124595
Paragorgia_arborea_NIWA_42001_JX124574
Paragorgia_arborea_NIWA_44609_JX124534
Paragorgia_arborea_NIWA_28425_JX124550
Paragorgia_arborea_NIWA_46377_JX124549
Paragorgia_arborea_NIWA_28160_JX124523
Paragorgia_arborea_NIWA_46315_JX124569
Paragorgia_arborea_NIWA_76238_JX124533
Paragorgia_arborea_USNM_1014919_JX124578_Sanchez2005
Paragorgia_arborea_RBCM_1601_EXAMINED
Paragorgia_arborea_USNM_1075745_JX124602
Paragorgia_arborea_USNM_1075760_JX124530
Paragorgia_arborea_AAndrews_DAVI3_JX124542
Paragorgia_arborea_USNM_56389
Paragorgia_arborea_USNM_1075766_JX124541
Paragorgia_arborea_USNM_1016320_JX124581
Paragorgia_arborea_USNM_1075753_JX124527
Paragorgia_arborea_BStone_20070178B01_JX124587
Paragorgia_arborea_USNM_1075761_JX124576
Paragorgia_arborea_USNM_1122233_JX124586
Paragorgia_arborea_USNM_1075754_JX124593
Paragorgia_arborea_USNM_1122237_JX124590
Paragorgia_arborea_USNM_1075746_JX124540
Paragorgia_arborea_USNM_1122240_JX124522
Paragorgia_arborea_AAndrews_DAVI1_JX124577
Paragorgia_arborea_USNM_1123932_JX124524
Paragorgia_arborea_USNM_1075744_JX124565
Paragorgia_arborea_USNM_1007340_JX124556_EXAMINED
Paragorgia_arborea_RBCM_0672_EXAMINED
Paragorgia_arborea_USNM_1075738_JX124526
Paragorgia_arborea_USNM_1027060_JX124560
Paragorgia_arborea_USNM_4091_JX124579
Paragorgia_arborea_MNHN_422_JX124564
Paragorgia_arborea_NIWA_41780_JX124558_EXAMINED
Paragorgia_arborea_USNM_1092764_JX124589
Paragorgia_arborea_NIWA_28123_JX124546
Paragorgia_arborea_NIWA_46320_JX124570
Paragorgia_arborea_USNM_80937_GQ293312
Paragorgia_arborea_NIWA_41854_JX124611
Paragorgia_arborea_NIWA_41999_JX124588
Paragorgia_arborea_USNM_1120444_JX124547
Paragorgia_arborea_USNM_4242_JX124567
Paragorgia_arborea_USNM_1092766_JX124554
Paragorgia_arborea_NIWA_28422_JX124601
Paragorgia_arborea_NIWA_3309_JX124535_Sanchez2005
Paragorgia_arborea_NIWA_3308_JX124552_Sanchez2005
Paragorgia_arborea_NIWA_46316_JX124573
Paragorgia_arborea_NIWA_3310_GQ293311_Sanchez2005
Paragorgia_arborea_USNM_80936_JX124543
Paragorgia_arborea_NIWA_28156_JX124610
Paragorgia_arborea_NIWA_28154_JX124553

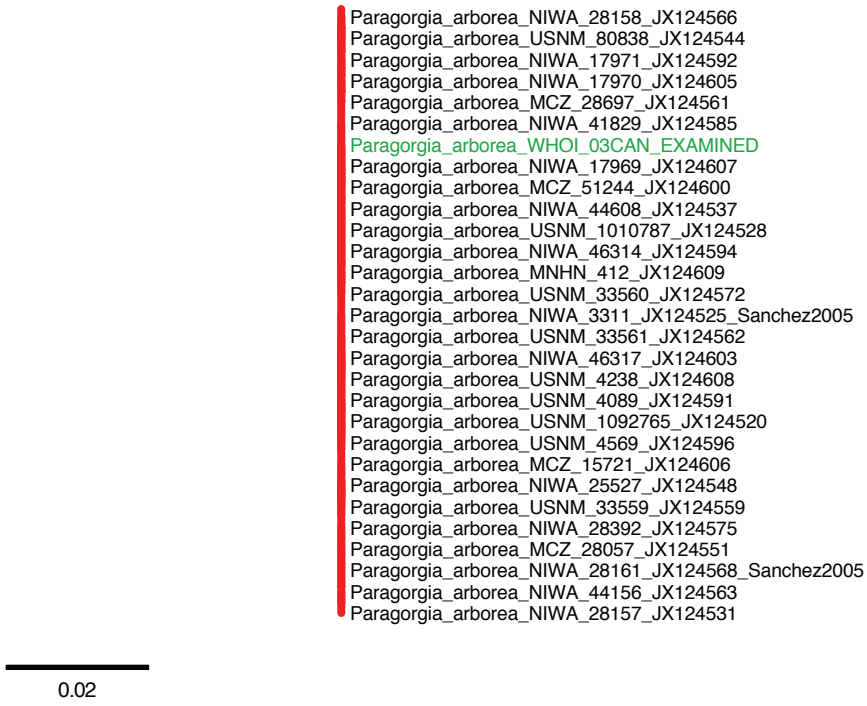
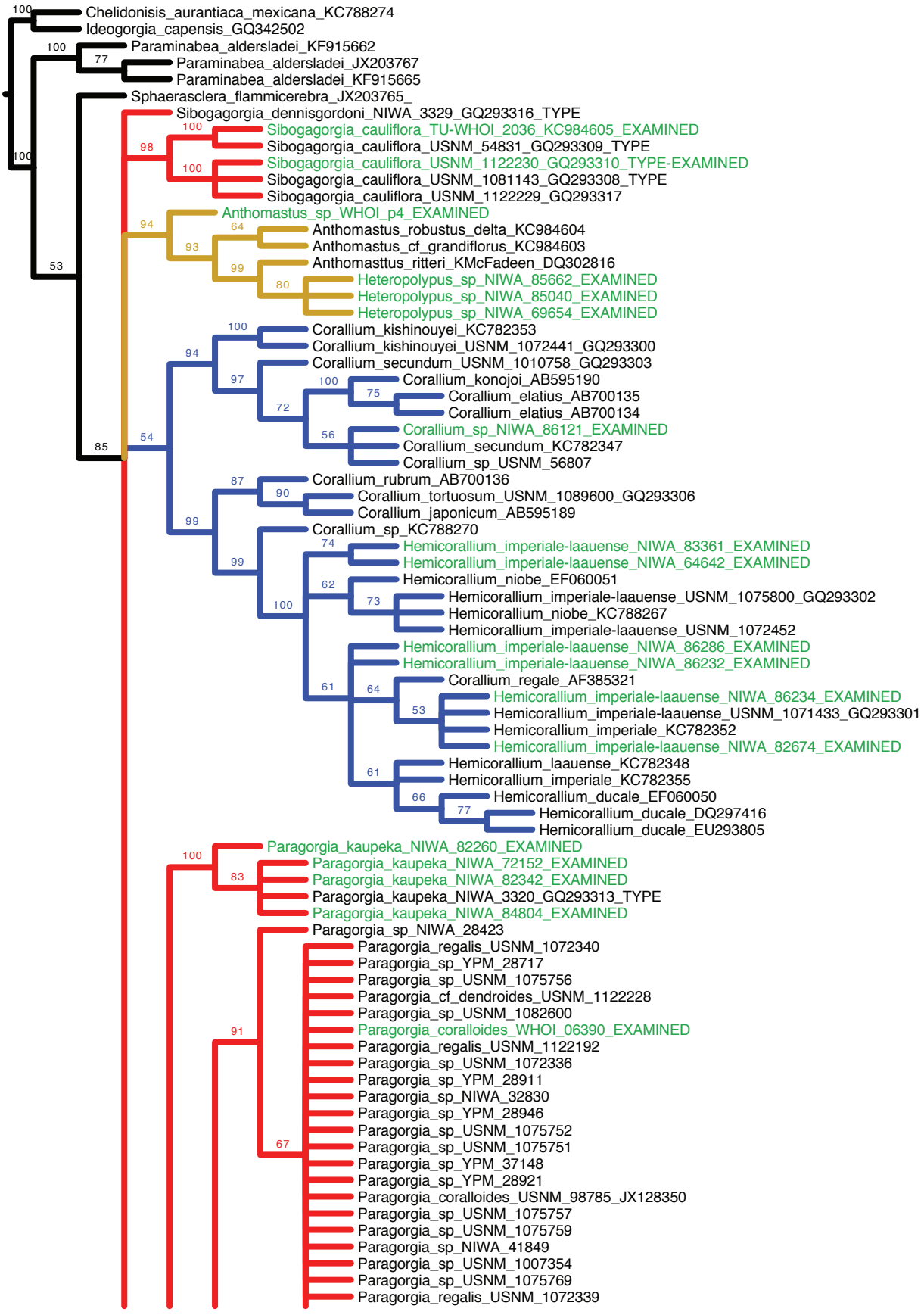
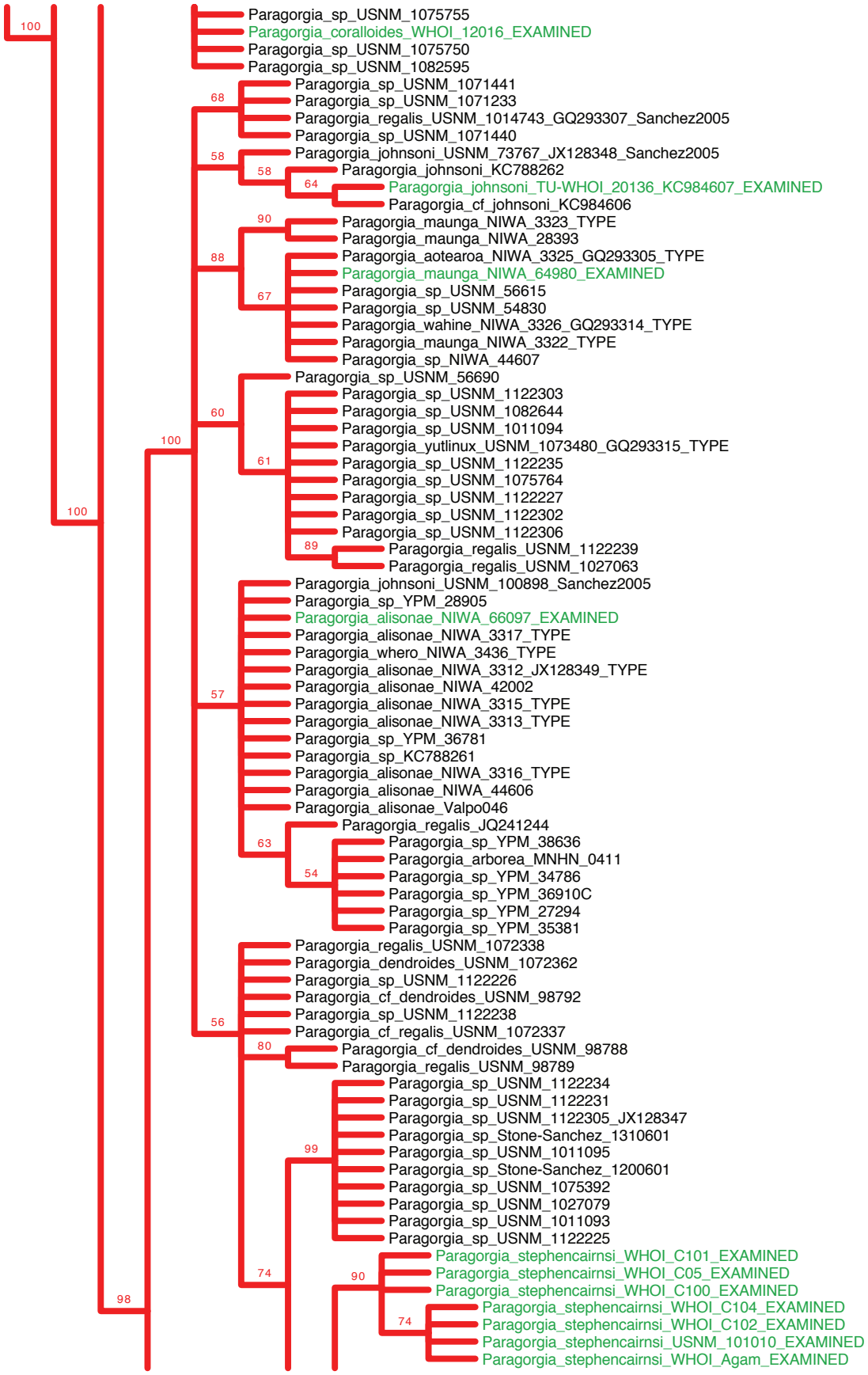
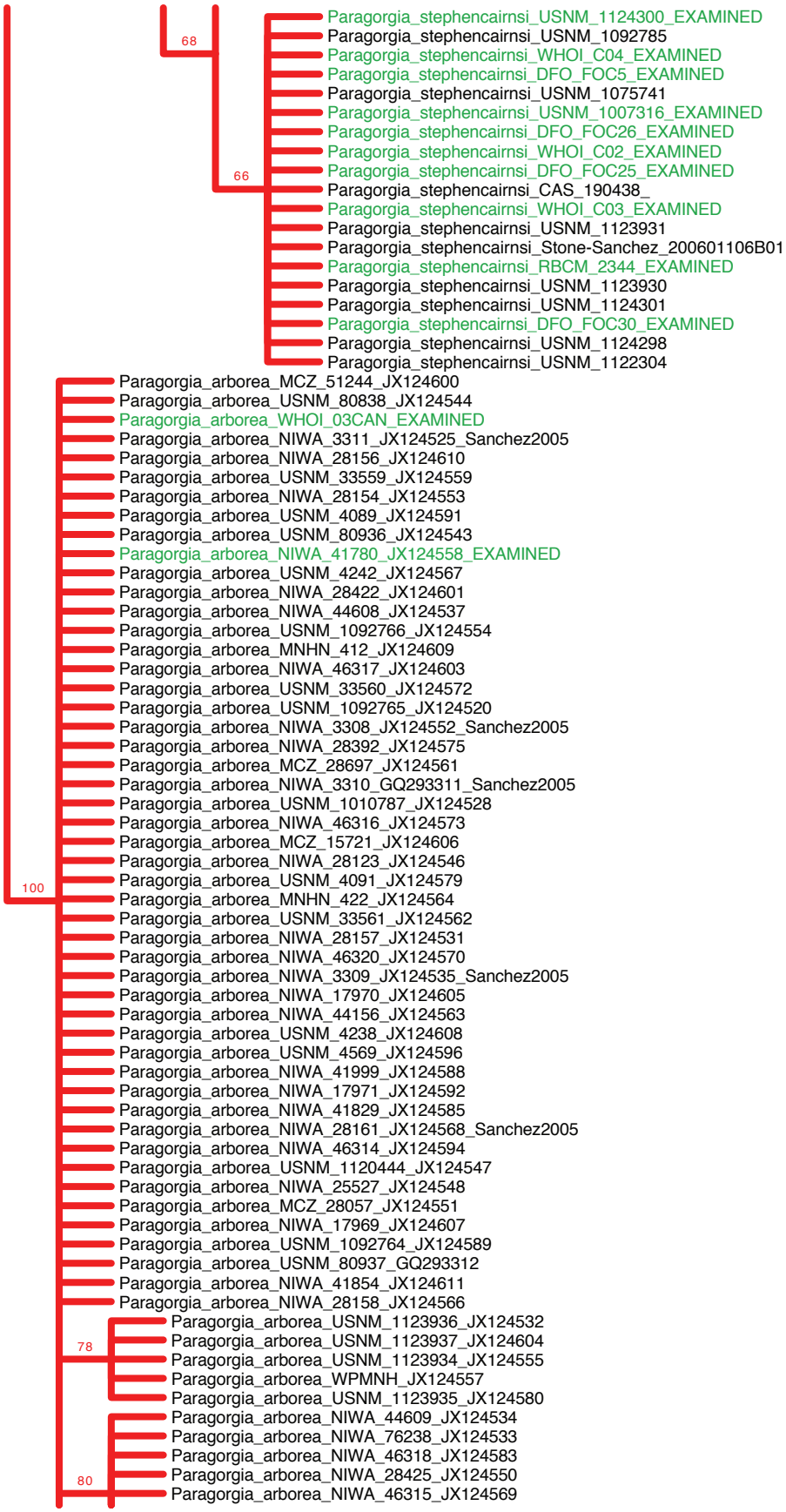


Figure S2. Mitochondrial *mtMutS* gene tree of all available sequences for the clade AC. Tree inferred from *mtMutS* sequence data from specimens examined in this study, GenBank, and additional specimens. Each family is indicated with a different branch color: red for Paragorgiidae; blue for Coralliidae; and yellow for Alcyoniidae. Outgroups are indicated with black branches. Specimens examined in detail in this study are indicated with green labels. Type specimens are labeled TYPE. Specimens examined in Sanchez (2005) are labeled "Sanchez2005". Tree was inferred with RAxML. Scale bars indicate substitutions per site.


Chelidonisis_aurantiaca_mexicana_KC788274
Ideogorgia_capensis_GQ342502
Paraminabea_aldersladei_KF915662
Paraminabea_aldersladei_JX203767
Paraminabea_aldersladei_KF915665
Sphaerasclera_flammicerebra_JX203765_
Sibogagorgia_dennisgordoni_NIWA_3329_GQ293316_TYPE
Sibogagorgia_cauliflora_TU-WHOI_2036_KC984605_EXAMINED
Sibogagorgia_cauliflora_USNM_54831_GQ293309_TYPE
Sibogagorgia_cauliflora_USNM_1122230_GQ293310_TYPE-EXAMINED
Sibogagorgia_cauliflora_USNM_1081143_GQ293308_TYPE
Sibogagorgia_cauliflora_USNM_1122229_GQ293317
Anthomastus_sp_WHOI_p4_EXAMINED
Anthomastus_robustus_delta_KC984604
Anthomastus_cf_grandiflorus_KC984603
Anthomasttus_ritteri_KMcFadeen_DQ302816
Heteropolypus_sp_NIWA_85662_EXAMINED
Heteropolypus_sp_NIWA_85040_EXAMINED
Heteropolypus_sp_NIWA_69654_EXAMINED
Corallium_kishinouyei_KC782353
Corallium_kishinouyei_USNM_1072441_GQ293300
Corallium_secundum_USNM_1010758_GQ293303
Corallium_konojoi_AB595190
Corallium_elatius_AB700135
Corallium_elatius_AB700134
Corallium_sp_NIWA_86121_EXAMINED
Corallium_secundum_KC782347
Corallium_sp_USNM_56807
Corallium_rubrum_AB700136
Corallium_tortuosum_USNM_1089600_GQ293306
Corallium_japonicum_AB595189
Corallium_sp_KC788270
Hemicorallium_imperiale-laauense_NIWA_83361_EXAMINED
Hemicorallium_imperiale-laauense_NIWA_64642_EXAMINED
Hemicorallium_niobe_EF060051
Hemicorallium_imperiale-laauense_USNM_1075800_GQ293302
Hemicorallium_niobe_KC788267
Hemicorallium_imperiale-laauense_USNM_1072452
Hemicorallium_imperiale-laauense_NIWA_86286_EXAMINED
Hemicorallium_imperiale-laauense_NIWA_86232_EXAMINED
Corallium_regale_AF385321
Hemicorallium_imperiale-laauense_NIWA_86234_EXAMINED
Hemicorallium_imperiale-laauense_USNM_1071433_GQ293301
Hemicorallium_imperiale_KC782352
Hemicorallium_imperiale-laauense_NIWA_82674_EXAMINED
Hemicorallium_laauense_KC782348
Hemicorallium_imperiale_KC782355
Hemicorallium_ducale_EF060050
Hemicorallium_ducale_DQ297416
Hemicorallium_ducale_EU293805
Paragorgia_kaupeka_NIWA_82260_EXAMINED
Paragorgia_kaupeka_NIWA_72152_EXAMINED
Paragorgia_kaupeka_NIWA_82342_EXAMINED
Paragorgia_kaupeka_NIWA_3320_GQ293313_TYPE
Paragorgia_kaupeka_NIWA_84804_EXAMINED
Paragorgia_sp_NIWA_28423
Paragorgia_regalis_USNM_1072340
Paragorgia_sp_YPM_28717
Paragorgia_sp_USNM_1075756
Paragorgia_cf_dendroides_USNM_1122228
Paragorgia_sp_USNM_1082600
Paragorgia_coralloides_WHOI_06390_EXAMINED
Paragorgia_regalis_USNM_1122192
Paragorgia_sp_USNM_1072336
Paragorgia_sp_YPM_28911
Paragorgia_sp_NIWA_32830
Paragorgia_sp_YPM_28946
Paragorgia_sp_USNM_1075752
Paragorgia_sp_USNM_1075751
Paragorgia_sp_YPM_37148
Paragorgia_sp_YPM_28921
Paragorgia_coralloides_USNM_98785_JX128350
Paragorgia_sp_USNM_1075757
Paragorgia_sp_USNM_1075759
Paragorgia_sp_NIWA_41849
Paragorgia_sp_USNM_1007354
Paragorgia_sp_USNM_1075769
Paragorgia_regalis_USNM_1072339


Paragorgia_sp_USNM_1075755
Paragorgia_coralloides_WHOI_12016_EXAMINED
Paragorgia_sp_USNM_1075750
Paragorgia_sp_USNM_1082595
Paragorgia_sp_USNM_1071441
Paragorgia_sp_USNM_1071233
Paragorgia_regalis_USNM_1014743_GQ293307_Sanchez2005
Paragorgia_sp_USNM_1071440
Paragorgia_johnsoni_USNM_73767_JX128348_Sanchez2005
Paragorgia_johnsoni_KC788262
Paragorgia_johnsoni_TU-WHOI_20136_KC984607_EXAMINED
Paragorgia_cf_johnsoni_KC984606
Paragorgia_maunga_NIWA_3323_TYPE
Paragorgia_maunga_NIWA_28393
Paragorgia_aotearoa_NIWA_3325_GQ293305_TYPE
Paragorgia_maunga_NIWA_64980_EXAMINED
Paragorgia_sp_USNM_56615
Paragorgia_sp_USNM_54830
Paragorgia_wahine_NIWA_3326_GQ293314_TYPE
Paragorgia_maunga_NIWA_3322_TYPE
Paragorgia_sp_NIWA_44607
Paragorgia_sp_USNM_56690
Paragorgia_sp_USNM_1122303
Paragorgia_sp_USNM_1082644
Paragorgia_sp_USNM_1011094
Paragorgia_yutlinux_USNM_1073480_GQ293315_TYPE
Paragorgia_sp_USNM_1122235
Paragorgia_sp_USNM_1075764
Paragorgia_sp_USNM_1122227
Paragorgia_sp_USNM_1122302
Paragorgia_sp_USNM_1122306
Paragorgia_regalis_USNM_1122239
Paragorgia_regalis_USNM_1027063
Paragorgia_johnsoni_USNM_100898_Sanchez2005
Paragorgia_sp_YPM_28905
Paragorgia_alisonae_NIWA_66097_EXAMINED
Paragorgia_alisonae_NIWA_3317_TYPE
Paragorgia_whero_NIWA_3436_TYPE
Paragorgia_alisonae_NIWA_3312_JX128349_TYPE
Paragorgia_alisonae_NIWA_42002
Paragorgia_alisonae_NIWA_3315_TYPE
Paragorgia_alisonae_NIWA_3313_TYPE
Paragorgia_sp_YPM_36781
Paragorgia_sp_KC788261
Paragorgia_alisonae_NIWA_3316_TYPE
Paragorgia_alisonae_NIWA_44606
Paragorgia_alisonae_Valpo046
Paragorgia_regalis_JQ241244
Paragorgia_sp_YPM_38636
Paragorgia_arborea_MNHN_0411
Paragorgia_sp_YPM_34786
Paragorgia_sp_YPM_36910C
Paragorgia_sp_YPM_27294
Paragorgia_sp_YPM_35381
Paragorgia_regalis_USNM_1072338
Paragorgia_dendroides_USNM_1072362
Paragorgia_sp_USNM_1122226
Paragorgia_cf_dendroides_USNM_98792
Paragorgia_sp_USNM_1122238
Paragorgia_cf_regalis_USNM_1072337
Paragorgia_cf_dendroides_USNM_98788
Paragorgia_regalis_USNM_98789
Paragorgia_sp_USNM_1122234
Paragorgia_sp_USNM_1122231
Paragorgia_sp_USNM_1122305_JX128347
Paragorgia_sp_Stone-Sanchez_1310601
Paragorgia_sp_USNM_1011095
Paragorgia_sp_Stone-Sanchez_1200601
Paragorgia_sp_USNM_1075392
Paragorgia_sp_USNM_1027079
Paragorgia_sp_USNM_1011093
Paragorgia_sp_USNM_1122225
Paragorgia_stephencairnsi_WHOI_C101_EXAMINED
Paragorgia_stephencairnsi_WHOI_C05_EXAMINED
Paragorgia_stephencairnsi_WHOI_C100_EXAMINED
Paragorgia_stephencairnsi_WHOI_C104_EXAMINED
Paragorgia_stephencairnsi_WHOI_C102_EXAMINED
Paragorgia_stephencairnsi_USNM_101010_EXAMINED
Paragorgia_stephencairnsi_WHOI_Agam_EXAMINED
100
68
58
58
64
90
88
67
60
100
61
100
89
57
63
54
56
80
99
74
90
74
98


Paragorgia_stephencairnsi_USNM_1124300_EXAMINED
Paragorgia_stephencairnsi_USNM_1092785
Paragorgia_stephencairnsi_WHOI_C04_EXAMINED
Paragorgia_stephencairnsi_DFO_FOC5_EXAMINED
Paragorgia_stephencairnsi_USNM_1075741
Paragorgia_stephencairnsi_USNM_1007316_EXAMINED
Paragorgia_stephencairnsi_DFO_FOC26_EXAMINED
Paragorgia_stephencairnsi_WHOI_C02_EXAMINED
Paragorgia_stephencairnsi_DFO_FOC25_EXAMINED
Paragorgia_stephencairnsi_CAS_190438_
Paragorgia_stephencairnsi_WHOI_C03_EXAMINED
Paragorgia_stephencairnsi_USNM_1123931
Paragorgia_stephencairnsi_Stone-Sanchez_200601106B01
Paragorgia_stephencairnsi_RBCM_2344_EXAMINED
Paragorgia_stephencairnsi_USNM_1123930
Paragorgia_stephencairnsi_USNM_1124301
Paragorgia_stephencairnsi_DFO_FOC30_EXAMINED
Paragorgia_stephencairnsi_USNM_1124298
Paragorgia_stephencairnsi_USNM_1122304
Paragorgia_arborea_MCZ_51244_JX124600
Paragorgia_arborea_USNM_80838_JX124544
Paragorgia_arborea_WHOI_03CAN_EXAMINED
Paragorgia_arborea_NIWA_3311_JX124525_Sanchez2005
Paragorgia_arborea_NIWA_28156_JX124610
Paragorgia_arborea_USNM_33559_JX124559
Paragorgia_arborea_NIWA_28154_JX124553
Paragorgia_arborea_USNM_4089_JX124591
Paragorgia_arborea_USNM_80936_JX124543
Paragorgia_arborea_NIWA_41780_JX124558_EXAMINED
Paragorgia_arborea_USNM_4242_JX124567
Paragorgia_arborea_NIWA_28422_JX124601
Paragorgia_arborea_NIWA_44608_JX124537
Paragorgia_arborea_USNM_1092766_JX124554
Paragorgia_arborea_MNHN_412_JX124609
Paragorgia_arborea_NIWA_46317_JX124603
Paragorgia_arborea_USNM_33560_JX124572
Paragorgia_arborea_USNM_1092765_JX124520
Paragorgia_arborea_NIWA_3308_JX124552_Sanchez2005
Paragorgia_arborea_NIWA_28392_JX124575
Paragorgia_arborea_MCZ_28697_JX124561
Paragorgia_arborea_NIWA_3310_GQ293311_Sanchez2005
Paragorgia_arborea_USNM_1010787_JX124528
Paragorgia_arborea_NIWA_46316_JX124573
Paragorgia_arborea_MCZ_15721_JX124606
Paragorgia_arborea_NIWA_28123_JX124546
Paragorgia_arborea_USNM_4091_JX124579
Paragorgia_arborea_MNHN_422_JX124564
Paragorgia_arborea_USNM_33561_JX124562
Paragorgia_arborea_NIWA_28157_JX124531
Paragorgia_arborea_NIWA_46320_JX124570
Paragorgia_arborea_NIWA_3309_JX124535_Sanchez2005
Paragorgia_arborea_NIWA_17970_JX124605
Paragorgia_arborea_NIWA_44156_JX124563
Paragorgia_arborea_USNM_4238_JX124608
Paragorgia_arborea_USNM_4569_JX124596
Paragorgia_arborea_NIWA_41999_JX124588
Paragorgia_arborea_NIWA_17971_JX124592
Paragorgia_arborea_NIWA_41829_JX124585
Paragorgia_arborea_NIWA_28161_JX124568_Sanchez2005
Paragorgia_arborea_NIWA_46314_JX124594
Paragorgia_arborea_USNM_1120444_JX124547
Paragorgia_arborea_NIWA_25527_JX124548
Paragorgia_arborea_MCZ_28057_JX124551
Paragorgia_arborea_NIWA_17969_JX124607
Paragorgia_arborea_USNM_1092764_JX124589
Paragorgia_arborea_USNM_80937_GQ293312
Paragorgia_arborea_NIWA_41854_JX124611
Paragorgia_arborea_NIWA_28158_JX124566
Paragorgia_arborea_USNM_1123936_JX124532
Paragorgia_arborea_USNM_1123937_JX124604
Paragorgia_arborea_USNM_1123934_JX124555
Paragorgia_arborea_WPMNH_JX124557
Paragorgia_arborea_USNM_1123935_JX124580
Paragorgia_arborea_NIWA_44609_JX124534
Paragorgia_arborea_NIWA_76238_JX124533
Paragorgia_arborea_NIWA_46318_JX124583
Paragorgia_arborea_NIWA_28425_JX124550
Paragorgia_arborea_NIWA_46315_JX124569
68
66
100
78
80

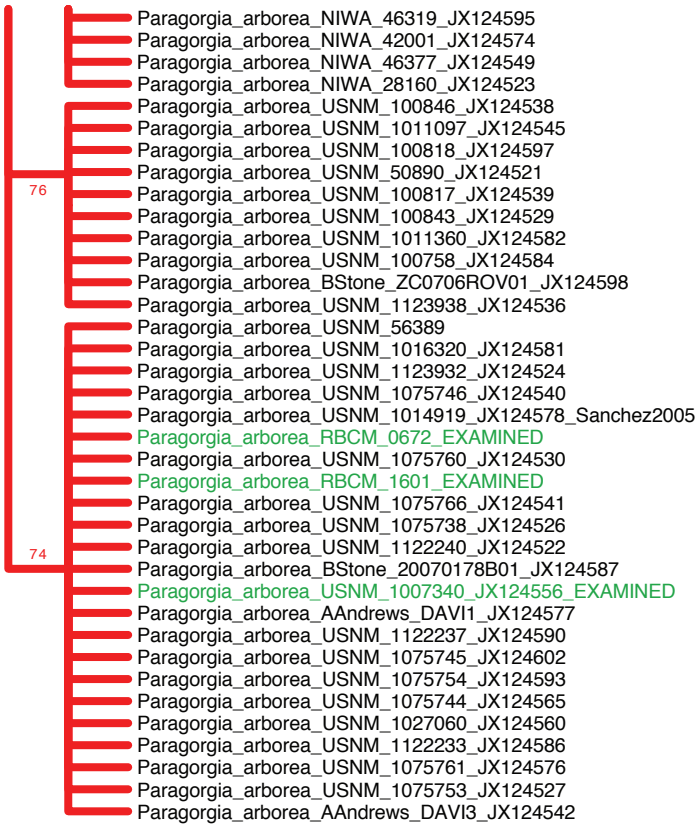


Figure S3. Mitochondrial *mtMutS* bootstrap support consensus tree of all available sequences for the clade AC. Tree inferred from *mtMutS* sequence data from specimens examined in this study, GenBank, and additional specimens. Each family is indicated with a different branch color: red for Paragorgiidae; blue for Coralliidae; and yellow for Alcyoniidae. Outgroups are indicated with black branches. Specimens examined in detail in this study are indicated with green labels. Type specimens are labeled "TYPE". Specimens examined in Sanchez (2005) are labeled "Sanchez2005". Tree was created with RAxML using a 50% majority consensus from 500 bootstrap replicates.

Table S1. Collection and sequence information for the specimens used in this study.

ID	Morphological Species	Delimitation model PABSTE	Collection	Date Collected	Locality	Depth (m)	Lat.	Lon.
p4	*Anthomastus sp*		WHOI	2011	Patagonian shelf			
86121	*Corallium sp*		NIWA	2012	Kermadec Ridge; Colville ridge volcano; TAN1213/18	380	-30.19	179.72
64642	*Hemicorallium imperiale-laauense*		NIWA	2010	Kermadec Ridge; Rumble II East seamount cone; TAN1007/97	1050	-35.42	178.65
83361	*Hemicorallium imperiale-laauense*		NIWA	2012	Bay of Plenty; Site SM1b, Matatara Knoll; TAN1206/168	948	-37.19	176.98
82674	*Hemicorallium imperiale-laauense*		NIWA	2012	Kermadec Ridge; Site SM2aa, summit of Whakatane Seamount; TAN1206/77	878	-36.81	177.47
86234	*Hemicorallium imperiale-laauense*		NIWA	2012	Kermadec Ridge; Northeast pimple volcano; TAN1213/22	483	-30.08	179.82
86286	*Hemicorallium imperiale-laauense*		NIWA	2012	Kermadec Ridge; Havre volcano; TAN1213/30	860	-31.13	-179.05
86232	*Hemicorallium imperiale-laauense*		NIWA	2012	Kermadec Ridge; Northeast pimple volcano; TAN1213/22	483	-30.08	179.82
69654	*Heteropolypus sp*		NIWA	2011	Chatam Rise; TRIP3306/78	495	-44.33	-177.22
85040	*Heteropolypus sp*		NIWA	2012	Chatam Rise; TAN1208/22	2098	-42.59	179.42
85662	*Heteropolypus sp*		NIWA	2012	Chatam Rise; TAN1208/61	1931	-42.59	179.59
66097	*Paragorgia alisonae*	*P. alisonae*	NIWA	2008	Cambell Plateau; TRIP2718/50	875	-50.02	175.00
41780	*Paragorgia arborea*	*P. arborea*	NIWA	2008	Chatam Rise; TRIP2617/120	600	-44.52	175.78
TC16_03	*Paragorgia arborea*	*P. arborea*	WHOI	2013	NE US Canyons; Munson Canyon; TowCam 16	540	40.54	-67.01
0672	*Paragorgia arborea*	*P. pacifica*	RBCM	2004	British Columbia; VE14280		53.33	-135.66
1007340	*Paragorgia arborea*	*P. pacifica*	USNM	2001	British Columbia; Vancouver Island	1168	48.44	-126.38
1601	*Paragorgia arborea*	*P. pacifica*	RBCM	2009	British Columbia; VE14444	695	53.31	-135.58
063902	*Paragorgia coralloides*	*P. coralloides*	WHOI	2005	New England Seamounts; Rehobot Seamount; H13	1821	37.46	-59.95
12016	*Paragorgia coralloides*	*P. coralloides*	WHOI	2003	New England Seamounts; Manning Seamount, station 4; AD3890	2000	38.23	-60.46
20136	*Paragorgia johnsoni*	*P. johnsoni*	TU/WHOI	2009	Gulf of Mexico; MC751; J2-464	438	28.19	-89.80
72152	*Paragorgia kaupeka*	*P. kaupeka*	NIWA	2011	Kermadec Ridge; Clark Seamount, chimney field, north cone; TAN1104/13	877	-36.45	177.84
82260	*Paragorgia kaupeka*	*P. kaupeka*	NIWA	2012	Kermadec Ridge; Site SM3a, summit of Clark Seamount; TAN1206/34	850	-36.45	177.84
82342	*Paragorgia kaupeka*	*P. kaupeka*	NIWA	2012	Kermadec Ridge; Site SM3a, Clark Seamount; TAN1206/40	1100	-36.45	177.84
84804	*Paragorgia kaupeka*	*P. kaupeka*	NIWA	2005	Kermadec Ridge; Clark Seamount; KOK0506/12	870	-36.45	177.84
64980	*Paragorgia maunga*	*P. maunga*	NIWA	2010	Kermadec Ridge; Silent II seamount; TAN1007/120	772	-35.17	178.89
1007316	*Paragorgia stephencairnsi*	*P. sp. nov*	USNM	2001	British Columbia; Vancouver Island	1168	48.44	-126.38
2344	*Paragorgia stephencairnsi*	*P. sp. nov*	RBCM	2004	British Columbia; VE13978	1194	53.37	-133.31

101010	*Paragorgia stephencairnsi*	*P. stephencairnsi*	USNM	2008	California; Piggy Bank, southern California; DW-026-02	283	33.92	-119.47
1124300	*Paragorgia stephencairnsi*	*P. stephencairnsi*	USNM	2006	British Columbia; Vancouver Island, Ohiat Island; OC 06/952	188	48.83	-125.13
Agam	*Paragorgia stephencairnsi*	*P. stephencairnsi*	WHOI	2012	British Columbia; Agamemnon Channel	32	49.72	-124.05
C02	*Paragorgia stephencairnsi*	*P. stephencairnsi*	WHOI	2013	British Columbia; Agamemnon Channel; Dive02	41	49.74	-124.03
C03	*Paragorgia stephencairnsi*	*P. stephencairnsi*	WHOI	2013	British Columbia; Agamemnon Channel; Dive02	41	49.74	-124.03
C04	*Paragorgia stephencairnsi*	*P. stephencairnsi*	WHOI	2013	British Columbia; Agamemnon Channel; Dive02	41	49.74	-124.03
C05	*Paragorgia stephencairnsi*	*P. stephencairnsi*	WHOI	2013	British Columbia; Agamemnon Channel; Dive02	41	49.74	-124.03
C100	*Paragorgia stephencairnsi*	*P. stephencairnsi*	WHOI	2013	British Columbia; Vancouver Island, Tahsis Inlet; Dive07	40	49.86	-126.67
C101	*Paragorgia stephencairnsi*	*P. stephencairnsi*	WHOI	2013	British Columbia; Vancouver Island, Tahsis Inlet; Dive07	40	49.86	-126.67
C102	*Paragorgia stephencairnsi*	*P. stephencairnsi*	WHOI	2013	British Columbia; Vancouver Island, Tahsis Inlet; Dive07	40	49.86	-126.67
C104	*Paragorgia stephencairnsi*	*P. stephencairnsi*	WHOI	2013	British Columbia; Vancouver Island, Tahsis Inlet; Dive07	40	49.86	-126.67
FOC25	*Paragorgia stephencairnsi*	*P. stephencairnsi*	DFO	2012	British Columbia; W of Graham Island; 2012-65	204	53.31	-133.03
FOC26	*Paragorgia stephencairnsi*	*P. stephencairnsi*	DFO	2012	British Columbia; W of Graham Island; 2012-65	221	53.30	-133.04
FOC30	*Paragorgia stephencairnsi*	*P. stephencairnsi*	DFO	2012	British Columbia; W of Graham Island; 2012-65	318	53.48	-133.07
FOC5	*Paragorgia stephencairnsi*	*P. stephencairnsi*	DFO	2009	British Columbia; E of Graham Island; 2009-47	201	52.13	-128.90
1122230	*Sibogagorgia cauliflora*		USNM	2006	California; Davidson seamount; dive 945	2502	35.83	-122.61
2036	*Sibogagorgia cauliflora*		TU/WHOI	2009	Gulf of Mexico; DC583; J2-454	2440	28.39	-87.39

Table S2. RAD sequencing results and filtering statistics.

				STACKS filtering		pyRAD filtering			
Morphological Species	**Delimitation model PABSTE**	**RAD-seq data file ID**	**Total sequenced reads**	**Reads discarded due to low quality**	**Retained reads**	**Reads that passed quality filtering**	**Trimmed reads due to detection of adapters**	**Total number of retained reads**	**% of retained reads after filtering steps**
Anthomastus sp		PoC_p4_ARG	2,207,834	151,391	1,312,903	1,076,607	38,570	1,115,177	50.51
Corallium sp		COR_86121_NZ	3,826,317	305,019	3,013,324	2,555,644	22,650	2,578,294	67.38
H. imperiale-laauense		COR_64642_NZ	4,344,702	239,625	4,105,077	3,697,256	21,823	3,719,079	85.60
H. imperiale-laauense		COR_83361_NZ	3,351,944	291,963	2,588,492	2,103,039	37,016	2,140,055	63.85
H. imperiale-laauense		HEM_82674_NZ	4,455,288	364,162	3,777,461	3,216,577	40,911	3,257,488	73.12
H. imperiale-laauense		HEM_86234_NZ	2,408,325	186,372	2,142,057	1,876,187	16,195	1,892,382	78.58
H. imperiale-laauense		HEM_86286_NZ	5,324,532	446,268	4,266,674	3,577,980	42,661	3,620,641	68.00
H. imperiale-laauense		PAR_86232_NZ	6,092,276	503,438	5,206,376	4,454,042	35,518	4,489,560	73.69
Heteropolypus sp		ANT_69654_NZ	4,300,289	349,226	3,645,814	3,139,309	32,395	3,171,704	73.76
Heteropolypus sp		ANT_85040_NZ	4,498,336	393,748	3,882,304	3,333,172	26,687	3,359,859	74.69
Heteropolypus sp		ANT_85662_NZ	3,487,138	283,856	2,812,750	2,382,184	40,135	2,422,319	69.46
P. alisonae	*P. alisonae*	PAR_66097_NZ	3,007,362	220,209	2,511,717	2,219,265	21,207	2,240,472	74.50
P. arborea	*P. arborea*	PAR_41780_NZ	6,668,080	367,144	6,300,936	5,655,328	33,758	5,689,086	85.32
P. arborea	*P. arborea*	PAR_TC16_03_CAN	2,259,880	167,106	1,799,244	1,556,363	17,448	1,573,811	69.64
P. arborea	*P. pacifica*	PAR_0672_BC	4,348,226	293,526	4,054,700	3,543,642	29,284	3,572,926	82.17
P. arborea	*P. pacifica*	PAR_1007340_BCD	2,808,882	251,102	2,431,829	2,063,940	25,831	2,089,771	74.40
P. arborea	*P. pacifica*	PAR_1601_BC	4,202,185	221,386	3,980,799	3,606,334	25,434	3,631,768	86.43
P. coralloides	*P. coralloides*	PAR_063902_NES	3,244,860	185,122	3,059,738	2,746,228	20,408	2,766,636	85.26
P. coralloides	*P. coralloides*	PAR_12016_NES	3,838,375	226,253	3,612,122	3,255,440	15,945	3,271,385	85.23
P. johnsoni	*P. johnsoni*	PAR_20136_GOM	7,751,624	422,203	7,329,421	6,623,446	38,692	6,662,138	85.95
P. kaupeka	*P. kaupeka*	PAR_72152_NZ	3,909,139	268,871	2,926,960	2,531,796	26,984	2,558,780	65.46
P. kaupeka	*P. kaupeka*	PAR_82260_NZ	5,511,873	475,292	4,558,291	3,824,683	42,782	3,867,465	70.17
P. kaupeka	*P. kaupeka*	PAR_82342_NZ	4,056,319	307,425	3,339,653	2,866,786	33,020	2,899,806	71.49
P. kaupeka	*P. kaupeka*	PAR_84804_NZ	2,828,879	230,493	2,313,997	1,968,495	19,602	1,988,097	70.28
P. maunga	*P. maunga*	PAR_11369_NZ	2,536,311	252,554	2,008,028	1,619,915	34,477	1,654,392	65.23
P. stephencairnsi	*P. sp. nov*	PAR_1007316_BCD	2,257,183	151,729	1,606,180	1,371,398	36,692	1,408,090	62.38
P. stephencairnsi	*P. sp. nov*	PAR_2344_BC	2,858,311	159,073	2,699,238	2,414,787	26,272	2,441,059	85.40
P. stephencairnsi	*P. stephencairnsi*	PAR_101010_CA	6,642,673	398,317	6,244,356	5,595,342	39,240	5,634,582	84.82
P. stephencairnsi	*P. stephencairnsi*	PAR_1124300_WA	3,843,250	215,859	3,627,391	3,260,864	27,921	3,288,785	85.57
P. stephencairnsi	*P. stephencairnsi*	PAR_Agam_BC	4,322,564	260,095	4,062,469	3,619,988	26,669	3,646,657	84.36

P. stephencairnsi	*P. stephencairnsi*	PAR_C02_BCS	2,149,938	153,978	1,594,818	1,377,002	16,157	1,393,159	64.80
P. stephencairnsi	*P. stephencairnsi*	PAR_C03_BCS	2,815,327	245,051	2,264,951	1,916,452	19,800	1,936,252	68.78
P. stephencairnsi	*P. stephencairnsi*	PAR_C04_BCS	3,240,713	267,042	2,754,373	2,379,284	22,261	2,401,545	74.11
P. stephencairnsi	*P. stephencairnsi*	PAR_C05_BCS	3,465,396	294,547	2,972,548	2,538,577	25,005	2,563,582	73.98
P. stephencairnsi	*P. stephencairnsi*	PAR_C100_BCS	5,998,914	480,465	4,984,766	4,330,930	38,445	4,369,375	72.84
P. stephencairnsi	*P. stephencairnsi*	PAR_C101_BCS	5,332,619	446,452	4,652,535	4,030,919	30,684	4,061,603	76.17
P. stephencairnsi	*P. stephencairnsi*	PAR_C102_BCS	4,347,757	348,662	3,710,381	3,221,934	26,559	3,248,493	74.72
P. stephencairnsi	*P. stephencairnsi*	PAR_C104_BCS	2,720,994	210,479	2,203,568	1,893,649	19,213	1,912,862	70.30
P. stephencairnsi	*P. stephencairnsi*	PAR_FOC25_BCD	5,479,109	453,625	4,693,804	4,073,503	34,583	4,108,086	74.98
P. stephencairnsi	*P. stephencairnsi*	PAR_FOC26_BCD	5,242,949	477,361	4,528,629	3,778,655	37,137	3,815,792	72.78
P. stephencairnsi	*P. stephencairnsi*	PAR_FOC30_BCD	4,417,520	337,501	3,781,973	3,312,791	32,621	3,345,412	75.73
P. stephencairnsi	*P. stephencairnsi*	PAR_FOC5_BCD	2,834,377	239,051	2,455,041	2,109,251	23,444	2,132,695	75.24
S. cauliflora		SIB_1122230_DAV	1,701,751	123,755	1,220,077	1,016,203	26,104	1,042,307	61.25
S. cauliflora		SIB_2036_GOM	2,638,533	191,392	2,373,232	2,077,092	20,727	2,097,819	79.51
		AVERAGE	3,944,974	292,232	3,395,704	2,950,279	28,840	2,979,119	74.27
		STD DEV.	1,392,573	105,332	1,352,276	1,219,146	7,968	1,222,799	8.09

Table S3. RAD-seq **backbone** clustering and matrix statistics.

				Number of loci recovered in final data set for each taxon															
Matrix	**Min. # of taxa per locus (m)**	**Cluster thres-hold (c)**	**Total # loci in matrix**	**ANT_85040_NZ**	**COR_86121_NZ**	**HEM_82674_NZ**	**PAR_0672_BC**	**PAR_11369_NZ**	**PAR_12016_NES**	**PAR_20136_GOM**	**PAR_66097_NZ**	**PAR_82260_NZ**	**PAR_PR27_BCD**	**PoC_p4_ARG**	**SIB_2036_GOM**	**Total # of variable sites (var)**	**Total # of phylogenetically informative sites (pis)**	**pis/var**	**% of missing data**
c80d5m4p3	4	80	62,726	10,588	24,303	25,369	41,190	44,891	38,333	46,858	42,472	41,810	45,377	7,154	26,243	601,763	109,290	0.18	49.6
c80d5m6p3	6	80	35,340	8,518	18,512	19,033	28,087	29,555	27,273	30,193	28,098	28,763	29,647	5,665	19,485	388,349	81,574	0.21	38.3
c80d5m9p3	9	80	10,333	5,176	8,607	8,701	9,558	9,778	9,528	9,879	9,477	9,635	9,794	3,458	8,183	132,803	32,017	0.24	20.7
c85d5m4p3	4	85	62,318	8,692	22,647	23,757	40,930	45,124	37,243	46,984	42,725	41,154	45,540	5,918	24,943	519,766	89,806	0.17	50.1
c85d5m6p3	6	85	33,785	7,020	17,146	17,608	27,047	28,504	26,012	29,192	27,197	27,640	28,608	4,715	18,225	328,464	66,419	0.20	38.3
c85d5m9p3	9	85	9,411	4,374	7,947	8,024	8,752	8,922	8,729	9,041	8,666	8,852	8,953	2,960	7,447	109,189	26,048	0.24	20.1
c90d5m4p3	4	90	58,765	5,796	18,775	19,255	37,994	43,101	33,306	44,722	40,899	37,911	43,075	4,073	21,371	402,245	63,938	0.16	51.4
c90d5m6p3	6	90	28,923	4,674	13,826	13,932	23,212	24,715	22,141	25,263	23,694	23,804	24,705	3,226	14,961	236,051	45,285	0.19	38.5
c90d5m9p3	9	90	7,312	2,943	6,277	6,318	6,855	6,955	6,836	7,048	6,779	6,905	6,984	2,029	5,815	72,149	16,695	0.23	19.4

Table S4. RAD-seq **PHYLO** clustering and matrix statistics.

Morphological Species	Delimitation model PABSTE	RAD-seq data file ID	Total number of clusters	Mean cluster depth	Std. dev. of cluster depth	Number of clusters with depth greater than 5	Mean depth of clusters with depth greater than 5	Std. dev. of cluster depth for clusters with depth greater than 5	Number of loci	Number of loci with >5 depth coverage	Number of loci with >5 depth and passed paralog filter	Number of sites across loci	Number of polymorphic sites	Frequency of polymorphic sites
Anthomastus sp		PoC_p4_ARG	142,551	5.7	11.5	54,060	11.9	16.9	142,551	54,050	46,191	3,917,645	12,866	0.0032841
Corallium sp		COR_86121_NZ	176,085	9.6	34.3	85,869	17.8	47.8	176,085	85,816	74,563	6,332,203	29,627	0.0046788
H. imperiale-laauense		COR_64642_NZ	177,897	4.3	26.1	29,162	16.7	62.8	177,897	29,125	14,235	1,205,608	6,515	0.0054039
H. imperiale-laauense		COR_83361_NZ	190,447	7.2	63.8	74,091	15.7	101.7	190,447	74,069	63,941	5,426,154	20,057	0.0036964
H. imperiale-laauense		HEM_82674_NZ	219,724	8.9	53.4	92,866	18.8	81.1	219,724	92,810	80,468	6,834,388	29,969	0.0043850
H. imperiale-laauense		HEM_86234_NZ	155,795	7.8	35.4	70,304	14.7	51.8	155,795	70,274	57,724	4,900,969	25,071	0.0051155
H. imperiale-laauense		HEM_86286_NZ	252,395	9.0	76.6	96,846	20.8	122.8	252,395	96,772	81,349	6,912,750	30,928	0.0044741
H. imperiale-laauense		PAR_86232_NZ	219,190	12.6	95.8	106,265	24.0	136.7	219,190	106,145	85,667	7,282,625	32,347	0.0044417
Heteropolypus sp		ANT_69654_NZ	151,154	15.1	52.2	88,266	24.7	66.7	151,154	88,178	80,757	6,864,699	18,791	0.0027373
Heteropolypus sp		ANT_85040_NZ	159,357	15.1	42.4	92,201	24.8	53.7	159,357	92,099	83,695	7,115,901	27,473	0.0038608
Heteropolypus sp		ANT_85662_NZ	165,100	10.5	120.2	82,840	19.2	169.2	165,100	82,793	74,272	6,309,974	23,863	0.0037818
P. alisonae	*P. alisonae*	PAR_66097_NZ	134,879	11.1	54.8	79,570	17.5	70.7	134,879	79,531	70,095	5,954,159	23,831	0.0040024
P. arborea	*P. arborea*	PAR_41780_NZ	107,360	17.7	78.9	76,515	24.0	92.7	107,360	76,367	68,928	5,855,658	20,899	0.0035690
P. arborea	*P. arborea*	PAR_TC16_03_CAN	138,933	7.7	46.4	62,427	14.9	68.6	138,933	62,399	54,495	4,625,417	13,801	0.0029837
P. arborea	*P. pacifica*	PAR_0672_BC	116,545	21.4	71.8	85,334	28.5	82.8	116,545	85,165	77,189	6,559,823	25,068	0.0038214
P. arborea	*P. pacifica*	PAR_1007340_BCD	149,606	9.9	30.4	76,311	17.6	41.0	149,606	76,259	67,539	5,736,203	22,134	0.0038587
P. arborea	*P. pacifica*	PAR_1601_BC	118,243	21.7	65.9	85,263	29.3	76.2	118,243	85,074	77,035	6,547,638	24,879	0.0037997
P. coralloides	*P. coralloides*	PAR_063902_NES	132,778	20.1	61.7	87,919	29.4	74.1	132,778	87,756	79,692	6,774,134	26,579	0.0039236
P. coralloides	*P. coralloides*	PAR_12016_NES	129,675	36.8	207.3	90,736	51.8	246.3	129,675	90,474	82,903	7,049,554	21,990	0.0031193
P. johnsoni	*P. johnsoni*	PAR_20136_GOM	122,432	29.8	487.2	89,665	40.0	568.9	122,432	89,630	81,301	6,911,399	23,857	0.0034518
P. kaupeka	*P. kaupeka*	PAR_72152_NZ	132,520	12.7	197.2	77,976	20.2	256.8	132,520	77,929	69,780	5,926,828	16,005	0.0027004
P. kaupeka	*P. kaupeka*	PAR_82260_NZ	186,768	13.1	77.9	89,892	25.4	110.9	186,768	89,766	80,556	6,845,659	17,710	0.0025870
P. kaupeka	*P. kaupeka*	PAR_82342_NZ	148,202	12.6	43.0	83,517	21.0	55.8	148,202	83,429	74,747	6,349,820	16,935	0.0026670
P. kaupeka	*P. kaupeka*	PAR_84804_NZ	124,756	10.5	119.9	69,174	17.3	160.6	124,756	69,119	61,022	5,182,009	15,300	0.0029525
P. maunga	*P. maunga*	PAR_11369_NZ	113,838	19.3	161.8	84,649	25.2	187.3	113,838	84,510	76,038	6,461,976	26,079	0.0040358
P. stephencairnsi	*P. sp nov*	PAR_1007316_BCD	166,722	5.9	17.7	60,649	12.9	27.9	166,722	60,636	51,368	4,357,427	19,584	0.0044944

P. stephencairnsi	*P. sp nov*	PAR_2344_BC	125,366	13.4	51.7	80,271	19.9	63.7	125,366	80,218	72,194	6,132,435	25,912	0.0042254
P. stephencairnsi	*P. stephencairnsi*	PAR_101010_CA	115,888	33.2	134.3	92,526	41.1	149.3	115,888	92,225	84,367	7,173,308	24,279	0.0033846
P. stephencairnsi	*P. stephencairnsi*	PAR_1124300_WA	109,466	20.9	103.4	83,749	26.8	117.6	109,466	83,645	76,157	6,471,962	21,532	0.0033270
P. stephencairnsi	*P. stephencairnsi*	PAR_Agam_BC	131,554	19.3	95.4	87,778	28.1	115.8	131,554	87,648	79,612	6,766,163	23,143	0.0034204
P. stephencairnsi	*P. stephencairnsi*	PAR_C02_BCS	111,985	9.0	22.7	63,618	14.2	29.0	111,985	63,597	56,380	4,786,614	16,361	0.0034181
P. stephencairnsi	*P. stephencairnsi*	PAR_C03_BCS	116,493	11.7	41.1	75,396	17.0	50.3	116,493	75,356	67,485	5,732,283	20,104	0.0035072
P. stephencairnsi	*P. stephencairnsi*	PAR_C04_BCS	128,452	13.0	47.9	81,845	19.4	59.1	128,452	81,792	73,789	6,269,830	21,010	0.0033510
P. stephencairnsi	*P. stephencairnsi*	PAR_C05_BCS	128,134	13.9	64.4	83,590	20.3	79.0	128,134	83,533	75,306	6,399,562	21,288	0.0033265
P. stephencairnsi	*P. stephencairnsi*	PAR_C100_BCS	152,749	19.9	55.9	92,858	31.6	69.2	152,749	92,693	84,309	7,168,326	22,907	0.0031956
P. stephencairnsi	*P. stephencairnsi*	PAR_C101_BCS	152,447	18.4	82.8	92,630	29.2	104.8	152,447	92,487	83,853	7,129,265	23,371	0.0032782
P. stephencairnsi	*P. stephencairnsi*	PAR_C102_BCS	131,680	17.3	46.6	88,133	24.9	55.3	131,680	88,038	79,983	6,798,340	21,816	0.0032090
P. stephencairnsi	*P. stephencairnsi*	PAR_C104_BCS	119,764	11.4	29.6	74,773	17.0	36.3	119,764	74,738	67,146	5,703,606	18,976	0.0033270
P. stephencairnsi	*P. stephencairnsi*	PAR_FOC25_BCD	154,768	18.3	56.3	93,419	29.2	70.4	154,768	93,284	84,570	7,190,050	24,801	0.0034494
P. stephencairnsi	*P. stephencairnsi*	PAR_FOC26_BCD	440,488	6.3	26.0	107,701	20.5	49.9	440,488	107,621	94,298	8,008,313	28,868	0.0036048
P. stephencairnsi	*P. stephencairnsi*	PAR_FOC30_BCD	314,181	7.6	44.3	95,360	21.1	78.8	314,181	95,276	84,774	7,202,768	24,900	0.0034570
P. stephencairnsi	*P. stephencairnsi*	PAR_FOC5_BCD	197,822	7.5	60.1	75,192	16.9	96.8	197,822	75,145	66,755	5,669,791	19,056	0.0033610
S. cauliflora		SIB_1122230_DAV	123,947	5.5	29.3	40,167	12.8	50.7	123,947	40,157	33,058	2,804,036	11,864	0.0042310
S. cauliflora		SIB_2036_GOM	127,391	11.3	36.4	71,374	18.6	47.3	127,391	71,315	63,094	5,358,510	17,312	0.0032307
		AVERAGE	157,171	13.9	76.4	80,744	22.6	99.0	157,171	80,658	71,425	6,068,313	21,810	0.0036393
		STANDARD DEV	59,920	7.2	77.0	15,153	8.0	89.6	59,920	15,119	14,833	1,262,657	5,318	0.0006234

Table S5. Nucleotide diversity and error rate estimates per specimen based on the **PHYLO** matrix

Morphological Species	PABSTE	RAD-seq data file ID	Nucleotide diversity (Pi)	Error Rate
Anthomastus sp		PoC_p4_ARG	0.01295412	0.0029032
Corallium sp		COR_86121_NZ	0.01467291	0.00258843
*H. imperiale-laauense**		COR_64642_NZ	0.04406216	0.01630744
H. imperiale-laauense		COR_83361_NZ	0.01399387	0.00269528
H. imperiale-laauense		HEM_82674_NZ	0.01440962	0.00268885
H. imperiale-laauense		HEM_86234_NZ	0.01676557	0.00254558
H. imperiale-laauense		HEM_86286_NZ	0.01526906	0.00234296
H. imperiale-laauense		PAR_86232_NZ	0.01633107	0.00217237
Heteropolypus sp		ANT_69654_NZ	0.00936378	0.00151989
Heteropolypus sp		ANT_85040_NZ	0.01122064	0.00146665
Heteropolypus sp		ANT_85662_NZ	0.01161827	0.00189506
P. alisonae	*P. alisonae*	PAR_66097_NZ	0.0127996	0.00205016
P. arborea	*P. arborea*	PAR_41780_NZ	0.01209999	0.00186767
P. arborea	*P. arborea*	PAR_TC16_03_CAN	0.01293451	0.00241858
P. arborea	*P. pacifica*	PAR_0672_BC	0.01189447	0.00170062
P. arborea	*P. pacifica*	PAR_1007340_BCD	0.01285038	0.00223148
P. arborea	*P. pacifica*	PAR_1601_BC	0.01205019	0.00164477
P. coralloides	*P. coralloides*	PAR_063902_NES	0.01182997	0.00146894
P. coralloides	*P. coralloides*	PAR_12016_NES	0.01035726	0.00116583
P. johnsoni	*P. johnsoni*	PAR_20136_GOM	0.01159218	0.00137491
P. kaupeka	*P. kaupeka*	PAR_72152_NZ	0.01085919	0.00209157
P. kaupeka	*P. kaupeka*	PAR_82260_NZ	0.01077562	0.00211998
P. kaupeka	*P. kaupeka*	PAR_82342_NZ	0.01082052	0.00225821
P. kaupeka	*P. kaupeka*	PAR_84804_NZ	0.01169178	0.00221668
P. maunga	*P. maunga*	PAR_11369_NZ	0.01225252	0.00176518
P. stephencairnsi	*P. sp. nov*	PAR_1007316_BCD	0.01488811	0.00295065
P. stephencairnsi	*P. sp. nov*	PAR_2344_BC	0.01260347	0.00188425
P. stephencairnsi	*P. stephencairnsi*	PAR_101010_CA	0.01063818	0.00134692
P. stephencairnsi	*P. stephencairnsi*	PAR_1124300_WA	0.01092481	0.00163497
P. stephencairnsi	*P. stephencairnsi*	PAR_Agam_BC	0.01117934	0.00151902
P. stephencairnsi	*P. stephencairnsi*	PAR_C02_BCS	0.01180863	0.00194975
P. stephencairnsi	*P. stephencairnsi*	PAR_C03_BCS	0.01129071	0.00175198
P. stephencairnsi	*P. stephencairnsi*	PAR_C04_BCS	0.01098867	0.00171446
P. stephencairnsi	*P. stephencairnsi*	PAR_C05_BCS	0.01117442	0.00167389
P. stephencairnsi	*P. stephencairnsi*	PAR_C100_BCS	0.01063313	0.00155401
P. stephencairnsi	*P. stephencairnsi*	PAR_C101_BCS	0.0109634	0.00160305
P. stephencairnsi	*P. stephencairnsi*	PAR_C102_BCS	0.01060609	0.00159815
P. stephencairnsi	*P. stephencairnsi*	PAR_C104_BCS	0.01124495	0.00175065
P. stephencairnsi	*P. stephencairnsi*	PAR_FOC25_BCD	0.01104859	0.00154584
P. stephencairnsi	*P. stephencairnsi*	PAR_FOC26_BCD	0.01263196	0.00180015
P. stephencairnsi	*P. stephencairnsi*	PAR_FOC30_BCD	0.01196811	0.00182162
P. stephencairnsi	*P. stephencairnsi*	PAR_FOC5_BCD	0.0119129	0.00186229
S. cauliflora		SIB_1122230_DAV	0.01525699	0.00328691
S. cauliflora		SIB_2036_GOM	0.01214983	0.00221517
		AVERAGE	0.01291599	0.00228532
		SD	0.00513429	0.00223806

*Excluded from calculations due to low number of loci

Table S6. Nucleotide diversity and error rate estimates per species based on the **PHYLO** matrix

		Nucleeotide diversity (Pi)			Per-site sequence error rate (Epsilon)		
Morphological Species	**Delimitation model PABSTE**	**mean**	**SD**	**SE**	**mean**	**SD**	**SE**
Anthomastus sp		0.01295412			0.00290320		
Corallium sp		0.01467291			0.00258843		
H. imperiale-laauense		0.01935775	0.01094005	0.00489254	0.00447727	0.00522005	0.00233448
Heteropolypus sp		0.01073423	0.00120338	0.00069477	0.00162720	0.00023350	0.00013481
P. alisonae	*P. alisonae*	0.01279960			0.00205016		
P. arborea	*P. arborea*	0.01245029	0.00056582	0.00040010	0.00193813	0.00041487	0.00029336
P. arborea	*P. pacifica*	0.01073423	0.00120338	0.00069477	0.00162720	0.00023350	0.00013481
P. coralloides	*P. coralloides*	0.01245029	0.00056582	0.00040010	0.00193813	0.00041487	0.00029336
P. johnsoni	*P. johnsoni*	0.01159218			0.00137491		
P. kaupeka	*P. kaupeka*	0.01241323	0.00175196	0.00087598	0.00229768	0.00048923	0.00024462
P. maunga	*P. maunga*	0.01225252			0.00176518		
P. stephencairnsi	*P. sp. nov*	0.01290646	0.00280248	0.00198165	0.00229281	0.00093033	0.00065784
P. stephencairnsi	*P. stephencairnsi*	0.01137950	0.00066172	0.00016543	0.00169174	0.00016364	0.00004091
S. cauliflora		0.01370341	0.00219709	0.00155358	0.00275104	0.00075783	0.00053587

Table S7. Predictions of # of RAD-tags in octocorals using PstI. Data for *Nematostella vectensis* obtained from the U.S. Joint Genome Institute (JGI-DOE) database. Data for *Acropora digitifera, Hydra vulgaris*, and *Alantina moseri* obtained from the U.S. National Center for Biotechnology Information (NCBI) WGS database. Observed frequency of recognition sequences and calculated probability based on a trinucleotide genome composition model were generated following the methodology described by Herrera et al. (2014). Octocoral genome size ranges were obtained by Luisa Dueñas from gorgoniid octocorals through flow cytomery at the Universidad de los Andes, Bogota, Colombia. Abbreviation: restriction sites (RS).

Cnidarians with sequenced genomes

Species	Common name	C-value	Genome size (Mbp) Unambiguous	Observed frequency of PstI RS per bp	Prob. of PstI recognition site per bp based on trinucleotide model
Nematostella vectensis	Starlet Anemone		297.39	0.00016661	0.00019608
Acropora digitifera	Staghorn Coral		364.97	0.00021313	0.00022777
Hydra vulgaris	Hydra		1189.96	0.00010830	0.00010871
Alatina moseri	Sea Wasp		1544.15	0.00020617	0.00021637

Predictions of # of PstI RS in known octocoral genome size range

Octocoral C-value	Octocoral Genome size (Mbp)	Predicted # of PstI RS based on *N. vectensis* observed frequency	Predicted # of PstI RS based on *A. digitifera* observed frequency	Predicted # of PstI RS based on *H. vulgaris* observed frequency	Predicted # of PstI RS based on *A. moseri* observed frequency
0.3	293.40	48,882.95	62,533.09	31,774.37	60,489.13
0.5	489.00	81,471.59	104,221.81	52,957.29	100,815.22

Octocoral C-value	Octocoral Genome size (Mbp)	Predicted # of PstI RS based on *N. vectensis* trinucleotide genome composition probability	Predicted # of PstI RS based on *A. digitifera* trinucleotide genome composition probability	Predicted # of PstI RS based on *H. vulgaris* trinucleotide genome composition probability	Predicted # of PstI RS based on *A. moseri* trinucleotide genome composition probability
0.3	293.40	57,529.94	66,828.40	31,895.01	63,482.46
0.5	489.00	95,883.23	111,380.67	53,158.36	105,804.09

Predictions of # of PstI RAD-tags in known octocoral genome size range

Octocoral C-value	Octocoral Genome size (Mbp)	Predicted # of PstI RS based on *N. vectensis* observed frequency	Predicted # of PstI RS based on *A. digitifera* observed frequency	Predicted # of PstI RS based on *H. vulgaris* observed frequency	Predicted # of PstI RS based on *A. moseri* observed frequency
0.3	293.40	97,765.91	125,066.17	63,548.75	120,978.26
0.5	489.00	162,943.18	208,443.62	105,914.58	201,630.43

Octocoral C-value	Octocoral Genome size (Mbp)	Predicted # of PstI RS based on *N. vectensis* trinucleotide genome composition probability	Predicted # of PstI RS based on *A. digitifera* trinucleotide genome composition probability	Predicted # of PstI RS based on *H. vulgaris* trinucleotide genome composition probability	Predicted # of PstI RS based on *A. moseri* trinucleotide genome composition probability
0.3	293.40	115,059.87	133,656.80	63,790.03	126,964.91
0.5	489.00	191,766.45	222,761.34	106,316.71	211,608.19

Table S8. Collection information for all specimens in the clade AC with available mtMutS sequences

Species	ID	Genbank	Collection	Catalog Number	Date	Locality	Depth (m)	Lat.	Lon.	Taxonomic remarks
Anthomasttus ritteri		DQ302816	K. McFadeen		1998	Off Pebble Beach: Californa: USA	300	36.58	-122.10	
Anthomastus cf grandiflorus		KC984603								
Anthomastus robustus delta		KC984604								
Anthomastus sp	p4		WHOI	p4	2011	Patagonian shelf				
Chelidonisis aurantiaca mexicana		KC788274								
Corallium elatius		AB700134								
Corallium elatius		AB700135								
Corallium japonicum		AB595189								
Corallium kishinouyei		GQ293300	USNM	1072441	2003	Off Laysan Island: Hawaii: USA	1490	25.70	-171.45	
Corallium kishinouyei		KC782353								
Corallium konojoi		AB595190								
Corallium regale		AF385321								
Corallium rubrum		AB700136								
Corallium secundum		GQ293303	USNM	1010758	2001	Off Maui: Hawaii: USA	240	20.88	-156.73	
Corallium secundum		KC782347								
Corallium sp	86121		NIWA	86121	2012	Kermadec Ridge; Colville ridge volcano; TAN1213/18	380	-30.19	179.72	
Corallium sp	Coralliumsp56807		USNM	56807	1978	USA, Hawaii, Oahu Island, Makapuu Point	366	21.30	-157.53	
Corallium sp		KC788270								
Corallium tortuosum		GQ293306	USNM	1089600	2003	New Caledonia	470-621	-23.71	168.26	
Hemicorallium ducale		DQ297416								
Hemicorallium ducale		EF060050								
Hemicorallium ducale		EU293805								
Hemicorallium imperiale		KC782352								
Hemicorallium imperiale		KC782355								
Hemicorallium imperiale-laauense	64642		NIWA	64642	2010	Kermadec Ridge; Rumble II East seamount cone; TAN1007/97	1050	-35.42	178.65	
Hemicorallium imperiale-laauense	82674		NIWA	82674	2012	Kermadec Ridge; Site SM2aa, summit of Whakatane Seamoun	878	-36.81	177.47	
Hemicorallium imperiale-laauense	83361		NIWA	83361	2012	Bay of Plenty; Site SM1b, Matatara Knoll; TAN1206/168	948	-37.19	176.98	
Hemicorallium imperiale-laauense	86232		NIWA	86232	2012	Kermadec Ridge; Northeast pimple volcano; TAN1213/22	483	-30.08	179.82	
Hemicorallium imperiale-laauense	86234		NIWA	86234	2012	Kermadec Ridge; Northeast pimple volcano; TAN1213/22	483	-30.08	179.82	
Hemicorallium imperiale-laauense	86286		NIWA	86286	2012	Kermadec Ridge; Havre volcano; TAN1213/30	860	-31.13	-179.05	
Hemicorallium imperiale-laauense	Claauense1072452		USNM	1072452	2003	USA, Hawaii, Laysan Island	1509	25.70	-171.44	
Hemicorallium imperiale-laauense		GQ293301	USNM	1071433	2004	Off Keahole Point: Hawaii Island: Hawaii: USA	867	19.80	-156.13	
Hemicorallium imperiale-laauense		GQ293302	USNM	1075800	2004	Pratt Seamount: Alaska: USA	1627	56.32	-142.44	
Hemicorallium laauense		KC782348								
Hemicorallium niobe		EF060051								
Hemicorallium niobe		KC788267								
Heteropolypus sp	69654		NIWA	69654	2011	Chatam Rise; TRIP3306/78	495	-44.33	-177.22	
Heteropolypus sp	85040		NIWA	85040	2012	Chatam Rise; TAN1208/22	2098	-42.59	179.42	
Heteropolypus sp	85662		NIWA	85662	2012	Chatam Rise; TAN1208/61	1931	-42.59	179.59	
Ideogorgia capensis		GQ342502								
Paragorgia alisonae	66097		NIWA	66097	2008	Cambell Plateau; TRIP2718/50	875	-50.02	175.00	
Paragorgia alisonae	Palisonae3312	JX128349	NIWA	3312	1998	New Zealand, Otara Hill, 1171/25, Z9596	980	-48.02	166.08	type

Paragorgia alisonae	Palisonae3313		NIWA	3313	1998	New Zealand, Otara Hill, 1171/25, Z9596	980	-48.02	166.08	type
Paragorgia alisonae	Palisonae3315		NIWA	3315	1998	New Zealand, Otara Hill, 1171/24, Z9595	940	-48.02	166.10	type
Paragorgia alisonae	Palisonae3316		NIWA	3316	1997	New Zealand, TAN9713/037, Z8981	1041	-44.96	174.19	type
Paragorgia alisonae	Palisonae3317		NIWA	3317	1998	New Zealand, TRIP1171/12, Z9583	935	-48.03	166.10	type
Paragorgia alisonae	Palisonae42002		NIWA	42002	2008	New Zealand, TRIP2551/258	930.0	-44.7850	-176.5833333333	
Paragorgia alisonae	Palisonae44606		NIWA	44606	2007	New Zealand, TRIP2521/9	1068	-44.7433	-177.0533333	
Paragorgia alisonae	ParagospValpo046			Valpo046		Chile, Valparaiso, deep seamount				
Paragorgia aotearoa		GQ293305	NIWA	3325	1996	New Zealand, Mt. Muck (Box Hill Complex), TAN9609/40	700	-42.83	176.92	type
Paragorgia arborea	0672		RBCM	011-00067-002	2004	British Columbia; VE14280		53.33	-135.66	
Paragorgia arborea	1601		RBCM	011-00160-001	2009	British Columbia; VE14444	695	53.31	-135.58	
Paragorgia arborea	41780	JX124558	NIWA	41870	2008	New Zealand, TRIP2617/120	600	-44.52	175.78	
Paragorgia arborea	1007340	JX124556	USNM	1007340	2001	British Columbia; Vancouver Island	1168	48.44	-126.38	
Paragorgia arborea	Paragosp1014919	JX124578	USNM	1014919	2003	Davidson Seamount, California, USA	1313	35.70	-122.70	Sanchez 2005
Paragorgia arborea	Paragosp1027060	JX124560	USNM	1027060	2003	Pioneer Seamount, South of farallon Islands, California, USA	1712	37.40	-123.44	
Paragorgia arborea	Paragosp1075738	JX124526	USNM	1075738	2004	Dickins Seamount, Gulf of Alaska, USA	760	54.55	-136.84	
Paragorgia arborea	Paragosp1075744	JX124565	USNM	1075744	2004	Dickins Seamount, Gulf of Alaska, USA	851	54.51	-136.91	
Paragorgia arborea	Paragosp1075745	JX124602	USNM	1075745	2004	Dickins Seamount, Gulf of Alaska, USA	849	54.51	-136.91	
Paragorgia arborea	Paragosp1075746	JX124540	USNM	1075746	2004	Welker Seamount, Gulf of Alaska, USA	780	55.05	-140.31	
Paragorgia arborea	Paragosp1075753	JX124527	USNM	1075753	2004	Welker Seamount, Gulf of Alaska, USA	1112	55.07	-140.41	
Paragorgia arborea	Paragosp1075754	JX124593	USNM	1075754	2004	Welker Seamount. Gulf of Alaska, USA	1084	55.07	-140.41	
Paragorgia arborea	Paragosp1075760	JX124530	USNM	1075760	2004	Pratt Seamount, Gulf of Alaska, USA	959	56.17	-142.70	
Paragorgia arborea	Paragosp1075761	JX124576	USNM	1075761	2004	Pratt Seamount, Gulf of Alaska, USA	941	56.17	-142.70	
Paragorgia arborea	Paragosp1075766	JX124541	USNM	1075766	2004	Welker Seamount, Gulf of Alaska, USA	1114	55.07	-140.41	
Paragorgia arborea	Paragosp17971	JX124592	NIWA	17971	2004	New Zealand, 1172/06, Z9566	1235	-44.80	-177.12	
Paragorgia arborea	Paragosp200701	JX124587	B. Stone	20070178B01	2004	Gulf of Alaska, USA	867	55.91	-154.02	
Paragorgia arborea	Paragosp44156	JX124563	NIWA	44156	2007	New Zealand, TRIP2416/54	720-741	-47.47	177.02	
Paragorgia arborea	Paragosp46314	JX124594	NIWA	46314	2008	New Zealand, TRIP2324/48	843-998	-50.05	174.73	
Paragorgia arborea	Paragosp46315	JX124569	NIWA	46315	2008	New Zealand, TRIP2571/65	888-101	-47.55	177.86	
Paragorgia arborea	Paragosp46316	JX124573	NIWA	46316	2007	New Zealand, TRIP2617/120	600	-44.52	175.77	
Paragorgia arborea	Paragosp46317	JX124603	NIWA	46317	2008	New Zealand, TRIP2494/13	931-102	-47.58	177.78	
Paragorgia arborea	Paragosp46318	JX124583	NIWA	46318	2008	New Zealand, TRIP2551/254	794-987	-44.73	-177.04	
Paragorgia arborea	Paragosp46319	JX124595	NIWA	46319	2006	New Zealand, TRIP2614		-49.50	176.00	
Paragorgia arborea	Paragosp46377	JX124549	NIWA	46377	1927	New Zealand, TRIP2571/53	952-1118	-50.00	176.06	
Paragorgia arborea	Paragosp56389		USNM	56389	1962	USA, Hawaii, Bushnell Seamount	1920	18.55	-155.44	
Paragorgia arborea	ParagospDAVI1	JX124577	A. Andre	DAVI1	2002	Davidson Seamount, California, USA	1313	35.75	-122.70	
Paragorgia arborea	ParagospDAVI3	JX124542	A. Andre	DAVI3	2004	Davidson Seamount, California, USA	1313	35.75	-122.70	
Paragorgia arborea	Paragospnizinski	JX124547	USNM	1120444	2008	off Maryland, USA	400	37.06	-74.62	
Paragorgia arborea	Parborea100758	JX124584	USNM	100758	1994	Aleutian Islands		52.00	-170.00	
Paragorgia arborea	Parborea100817	JX124539	USNM	100817	1994	Atka Island, Andreanof Islands, Aleutian Islands		53.00	-174.00	
Paragorgia arborea	Parborea100818	JX124597	USNM	100818	1994	Semisopochnoi Island, Rat Islands, Aleutian Islands		52.17	179.72	
Paragorgia arborea	Parborea100843	JX124529	USNM	100843	1994	Tanaga Island, Andreanof Islands, Aleutian Islands		52.00	-178.00	
Paragorgia arborea	Parborea100846	JX124538	USNM	100846	1994	Yunaska Island, Islands of Four Mountains, Aleutian Islands		53.00	-171.00	
Paragorgia arborea	Parborea1010787	JX124528	USNM	1010787	2000	Norfolk Canyon, Virginia, USA	375-489	37.07	-74.66	
Paragorgia arborea	Parborea1011097	JX124545	USNM	1011097	2002	Buldir Reef, Rat Islands, Aleutian Islands	160	51.96	176.83	
Paragorgia arborea	Parborea1011360	JX124582	USNM	1011360	2001	off Umnak Island, Fox Islands, Aleutian Islands	102	53.68	-169.11	

Paragorgia arborea	Parborea1092764	JX124589	USNM	1092764	2000	East of Virginia Beach, Virginia, USA	375-489	37.07	-74.66	
Paragorgia arborea	Parborea1092765	JX124520	USNM	1092765	2000	East of Virginia Beach, Virginia, USA	375-489	37.07	-74.66	
Paragorgia arborea	Parborea1092766	JX124554	USNM	1092766	2000	East of Virginia Beach, Virginia, USA	375-489	37.07	-74.66	
Paragorgia arborea	Parborea17969	JX124607	NIWA	17969	2002	New Zealand, 1621/18, Z11010	900	-44.74	-177.19	
Paragorgia arborea	Parborea17970	JX124605	NIWA	17970	2001	New Zealand, AEX0101/80, Z10956	753	-44.74	-177.19	
Paragorgia arborea	Parborea25527	JX124548	NIWA	25527	2004	New Zealand, TAN0408/23	826	-42.83	177.42	
Paragorgia arborea	Parborea28123	JX124546	NIWA	28123	2000	New Zealand, 1390/12, Z11161	872	-47.31	165.83	
Paragorgia arborea	Parborea28154	JX124553	NIWA	28154	1981	New Zealand, T16	427	-43.35	178.66	
Paragorgia arborea	Parborea28156	JX124610	NIWA	28156	2002	New Zealand, 1621/08, Z11008	920	-33.92	167.92	
Paragorgia arborea	Parborea28157	JX124531	NIWA	28157	1999	New Zealand, TRIP1223/29	959	-44.58	-177.88	
Paragorgia arborea	Parborea28158	JX124566	NIWA	28158	2001	New Zealand, AEX0101/80, Z10907	753	-44.74	-177.18	
Paragorgia arborea	Parborea28160	JX124523	NIWA	28160	2001	New Zealand				
Paragorgia arborea	Parborea28161	JX124568	NIWA	28161	2001	New Zealand, AEX0101/80, Z10920	753	-44.74	-176.81	Sanchez 2005
Paragorgia arborea	Parborea28392	JX124575	NIWA	28392	2001	New Zealand, AEX0101/80, Z10956	753	-44.74	-177.19	
Paragorgia arborea	Parborea28422	JX124601	NIWA	28422	2001	New Zealand, AEX0101/80, Z10920	753	-44.74	-176.81	
Paragorgia arborea	Parborea28425	JX124550	NIWA	28425	1997	New Zealand, TAN9713/52, Z8979	858	-44.45	-179.96	
Paragorgia arborea	Parborea3308	JX124552	NIWA	3308	2002	New Zealand, Z10987	1225	-33.93	167.92	Sanchez 2005
Paragorgia arborea	Parborea3309	JX124535	NIWA	3309	2002	New Zealand, Z11009	955	-33.93	167.91	Sanchez 2005
Paragorgia arborea	Parborea3310	GQ293311	NIWA	3310	1999	New Zealand, Z9862	687	-44.75	174.82	Sanchez 2005
Paragorgia arborea	Parborea3311	JX124525	NIWA	3311	1996	New Zealand, Southern Havre trough, X700	1525	-35.84	177.91	Sanchez 2005
Paragorgia arborea	Parborea33559	JX124559	USNM	33559	1878	Fishing Banks, North Carolina, USA	457	36.00	-74.00	
Paragorgia arborea	Parborea33561	JX124562	USNM	33561		Off NE North America, USA				
Paragorgia arborea	Parborea4089	JX124591	USNM	4089		Sable Island, 50 Mile E Of E Light, Nova Scotia, Canada	512	43.90	-58.80	
Paragorgia arborea	Parborea4091	JX124579	USNM	4091	1879	Banquereau Bank, Nova Scotia, Canada	366	44.58	-57.68	
Paragorgia arborea	Parborea4178B	JX124524	USNM	1123932	2002	South of Trinity Islands, Aleutian Islands	746	55.87	-154.06	
Paragorgia arborea	Parborea41829	JX124585	NIWA	41829	2006	New Zealand, TRIP2324/76	1044	-47.25	178.33	
Paragorgia arborea	Parborea41854	JX124611	NIWA	41854	2007	New Zealand, TRIP2494/10	867-986	-47.53	177.87	
Paragorgia arborea	Parborea41999	JX124588	NIWA	41999	2007	New Zealand, TRIP2551/50	1203-12	-44.50	-174.79	
Paragorgia arborea	Parborea42001	JX124574	NIWA	42001	2007	New Zealand, TRIP2551/55	1283-13	-44.50	-174.82	
Paragorgia arborea	Parborea4238	JX124608	USNM	4238	1879	Banquereau Bank, South Of, Nova Scotia, Canada	457	43.90	-58.67	
Paragorgia arborea	Parborea4242	JX124567	USNM	4242	1879	Grand Banks, W Part Of, Newfoundland, Canada		45.00	-54.00	
Paragorgia arborea	Parborea44608	JX124537	NIWA	44608	2007	New Zealand, TRIP2506/81	1106-13	-46.91	171.88	
Paragorgia arborea	Parborea44609	JX124534	NIWA	44609	2007	New Zealand, TRIP2506/135	870-967	-47.53	177.92	
Paragorgia arborea	Parborea4569	JX124596	USNM	4569	1879	Sable Island Bank, Nova Scotia, Canada	457	43.42	-60.00	
Paragorgia arborea	Parborea46320	JX124570	NIWA	46320	2008	New Zealand, TRIP2320/70	750-855	-46.48	170.60	
Paragorgia arborea	Parborea50890	JX124521	USNM	50890	1927	Burdwood Bank, S Of Falkland Islands, Scotia Sea		-54.50	-59.10	
Paragorgia arborea	Parborea80838	JX124544	USNM	80838	1979	Baltimore Canyon, Off Eastern Shore, Maryland, USA	480	38.17	-73.84	
Paragorgia arborea	Parborea80936	JX124543	USNM	80936	1979	Lydonia Canyon, Massachusetts, USA	680-370	40.38	-67.66	
Paragorgia arborea	Parborea80937	GQ293312	USNM	80937	1979	Lydonia Canyon, Massachusetts, USA	613-430	40.38	-67.66	
Paragorgia arborea	ParboreaJ2095271	JX124532	USNM	1123936	2004	Amlia Island, Andreanof Islands, Aleutian Islands	843	51.81	-173.83	
Paragorgia arborea	ParboreaJ2095272	JX124580	USNM	1123935	2004	Amlia Island, Andreanof Islands, Aleutian Islands	843	51.81	-173.83	
Paragorgia arborea	ParboreaJ2099211	JX124604	USNM	1123937	2004	Adak Canyon, Andreanof Islands, Aleutian Islands	1269	51.51	-177.04	
Paragorgia arborea	ParboreaJ210441	JX124536	USNM	1123938	2004	Amchitka Pass, Andreanof Islands, Aleutian Islands	747	51.72	-179.58	
Paragorgia arborea	ParboreaJ210462	JX124555	USNM	1123934	2004	Amchitka Pass, Andreanof Islands, Aleutian Islands	857	51.68	-179.58	
Paragorgia arborea	ParboreaMCZ1572	JX124606	MCZ	15721	2002	Atlantic Ocean	480	37.67	-74.65	

Paragorgia arborea	ParboreaMCZ2805	JX124551	MCZ	28057	2001	Nantucket Island, Massachussetts, USA	156	41.97	-65.87	
Paragorgia arborea	ParboreaMCZ2869	JX124561	MCZ	28697	1878	off Maine, USA	245	42.60	-65.73	
Paragorgia arborea	ParboreaMCZ5124	JX124600	MCZ	51244	1979	Georges Bank, Massachusetts, USA		41.00	-67.00	
Paragorgia arborea	ParboreaMNHN0411		MNHN	0411	1997	France, North Altlantic Ocean, France	700	48.83	-11.33	
Paragorgia arborea	ParboreaMNHN04	JX124609	MNHN	412		off Norway		66.70	11.60	
Paragorgia arborea	ParboreaMNHN04	JX124564	MNHN	422	2002	Trondhjems Fjord, Norway		63.50	10.50	
Paragorgia arborea	Parboreap33560	JX124572	USNM	33560		Fishing Banks, North Carolina, USA		45.00	-53.50	
Paragorgia arborea	ParboreaT661A10	JX124590	USNM	1122237	2002	Rodriguez Seamount, California, USA	894.5	34.06	-121.08	
Paragorgia arborea	ParboreaT662A28	JX124522	USNM	1122240	2002	San Juan Seamount, California, USA	1362.9	32.97	-121.04	
Paragorgia arborea	ParboreaT662A29	JX124586	USNM	1122233	2002	San Juan Seamount, California, USA	1360.8	32.97	-121.04	
Paragorgia arborea	ParboreaZ11166	JX124533	NIWA	76238	2007	New Zeland	891	-44.44	175.54	
Paragorgia arborea	ParboreaZC0706	JX124598	B. Stone	ZC0706ROV01		Zhemchung Canyon, Bering Sea	171	-47.54	177.93	
Paragorgia arborea	PcfarboreJapan	JX124557	WPMNH		2005	Off Yaizu-shi, Shijuoka Prof., Japan	760-800	33.00	138.40	
Paragorgia arborea	Ppacifica1016320	JX124581	USNM	1016320	2002	British Columbia, Canada	1152-115	53.70	-133.42	
Paragorgia arborea	TC16_03		WHOI	HB1302_TC016	2013	NE US Canyons; Munson Canyon; TowCam 16	540	40.54	-67.01	
Paragorgia cf dendroides	Pcfdendro98788		USNM	98788	1996	USA, Hawaii, Lanai Island, Keanapapa Point, SW of Point	1007	20.78	-157.15	
Paragorgia cf dendroides	Pcfdendro98792		USNM	98792	1996	USA, Hawaii, Lanai Island, Keanapapa Point, SW of Point	1018	20.78	-157.15	
Paragorgia cf dendroides	PcfdendroT662A30		USNM	1122228	200?	USA, California, San Juan Seamount	1237.6	32.97	-121.03	
Paragorgia cf johnsoni		KC984606								
Paragorgia cf regalis	Pcfregalis1072337		USNM	1072337	2003	USA, Hawaii, Pioneer Bank	1211	25.81	-173.50	
Paragorgia coralloides	06390		WHOI	REH112-5	2005	New England Seamounts; Rehobot Seamount; H13	1821	37.46	-59.95	
Paragorgia coralloides	12016		WHOI	leftover sieving	2003	New England Seamounts; Manning Seamount, station 4; AD38	2000	38.23	-60.46	
Paragorgia coralloides	Pcfcorallo98785	JX128350	USNM	98785	1995	East Pacific Rise, off Mexico	1950	12.73	-102.60	
Paragorgia dendroides	Pdendroid1072362		USNM	1072362	2003	USA, Hawaii, Necker Island, Seamount East of Island	1536	23.30	-163.70	
Paragorgia johnsoni	20136	KC984607	TU-WHC	0909-Oct1	2009	Gulf of Mexico; MC751; J2-464	438	28.19	-89.80	
Paragorgia johnsoni	Paragosp100898		USNM	100898	2000	USA, North Atlantic Ocean, South of Georges Bank		39.86	-67.42	Sanchez 2005
Paragorgia johnsoni	Pjohnsoni73767	JX128348	USNM	73767	1984	Little Bahama Bank, Bahamas	608	27.10	-79.70	Sanchez 2005
Paragorgia johnsoni		KC788262								
Paragorgia kaupeka	72152		NIWA	72152	2011	Kermadec Ridge; Clark Seamount, chimney field, north cone; ´	877	-36.45	177.84	
Paragorgia kaupeka	82260		NIWA	82260	2012	Kermadec Ridge; Site SM3a, summit of Clark Seamount; TAN	850	-36.45	177.84	
Paragorgia kaupeka	82342		NIWA	82342	2012	Kermadec Ridge; Site SM3a, Clark Seamount; TAN1206/40	1100	-36.45	177.84	
Paragorgia kaupeka	84804		NIWA	84804	2005	Kermadec Ridge; Clark Seamount; KOK0506/12	870	-36.45	177.84	
Paragorgia kaupeka	Pkaupeka3320	GQ293313	NIWA	3320	1989	New Zealand, X152	820	-36.16	176.81	type
Paragorgia maunga	64980		NIWA	64980	2010	Kermadec Ridge; Silent II seamount; TAN1007/120	772	-35.17	178.89	
Paragorgia maunga	Pmaunga28393		NIWA	28393	1999	New Zealand, Z9779	1121	-34.12	174.90	
Paragorgia maunga	Pmaunga3322		NIWA	3322	2002	New Zealand, Wanganella Bank, TRIP, Z10989	1082	-33.89	167.94	type
Paragorgia maunga	Pmaunga3323		NIWA	3323	1999	New Zealand, Z9779	1121	-34.12	174.90	type
Paragorgia regalis	ParagospT629A6		USNM	1122239	200?	USA, California, Rodriguez Seamount	1843.8	33.95	-121.14	
Paragorgia regalis	Pcfdendro98789		USNM	98789	1996	USA, Hawaii, Lanai Island, Keanapapa Point, SW of Point	1018	20.78	-157.15	
Paragorgia regalis	PdendroidT630A5		USNM	1122192	200?	USA, California, Rodriguez Seamount	1031.8	34.00	-121.10	
Paragorgia regalis	Pregalis1027063		USNM	1027063	2003	USA, California, Rodriguez Seamount, West of San Miguel Pa	1840	33.95	-121.14	
Paragorgia regalis	Pregalis1072338		USNM	1072338	2003	USA, Hawaii, Laysan Island, SE of Island	1136	25.67	-171.41	
Paragorgia regalis	Pregalis1072339		USNM	1072339	2003	USA, Hawaii, Pioneer Bank	1743.7	25.57	-173.51	
Paragorgia regalis	Pregalis1072340		USNM	1072340	2004	USA, Hawaii, Pioneer Bank	1744.7	25.57	-173.51	
Paragorgia regalis		GQ293307	USNM	1014743	2003	Cross Seamount: Hawaii: USA	452	19.74	-158.30	Sanchez 2005

Paragorgia regalis		JQ241244							
Paragorgia sp	Paragosp1007354		USNM	1007354	2001	Canada, British Columbia, Vancouver Island	1168	48.44	-126.38
Paragorgia sp	Paragosp1011093		USNM	1011093	2002	USA, Bering Sea, Forrester Island	601-800	55.40	-134.83
Paragorgia sp	Paragosp1011094		USNM	1011094	2002	USA, Bering Sea, East of Kodiak Island	601-836	58.22	-148.70
Paragorgia sp	Paragosp1011095		USNM	1011095	2002	USA, Bering Sea, Alaska, Alexander Archipielago, Baranof Isl	601-715	57.19	-136.24
Paragorgia sp	Paragosp1027079		USNM	1027079		USA, California, Rodriguez Seamount, West of San Miguel Pa	1030	34.05	-121.10
Paragorgia sp	Paragosp1071233		USNM	1071233	2004	USA, Hawaii, Cross Seamount	410	18.73	-158.26
Paragorgia sp	Paragosp1071440		USNM	1071440	2004	USA, Hawaii, Cross Seamount	427.51	18.73	-158.26
Paragorgia sp	Paragosp1071441		USNM	1071441	2004	USA, Hawaii, Cross Seamount	427.51	18.73	-158.26
Paragorgia sp	Paragosp1072336		USNM	1072336	2003	USA, Hawaii, Raita Bank	573	25.63	-169.32
Paragorgia sp	Paragosp1075392		USNM	1075392	2004	USA, Gulf of Alaska, Alaska, Giacomini Seamount	733	56.42	-146.37
Paragorgia sp	Paragosp1075750		USNM	1075750	2004	USA, Gulf of Alaska, Alaska, Welker Seamount	718	55.05	-140.31
Paragorgia sp	Paragosp1075751		USNM	1075751	2004	USA, Gulf of Alaska, Alaska, Welker Seamount	782	55.05	-140.31
Paragorgia sp	Paragosp1075752		USNM	1075752	2004	USA, Gulf of Alaska, Alaska, Welker Seamount	1119	55.06	-140.41
Paragorgia sp	Paragosp1075755		USNM	1075755	2004	USA, Gulf of Alaska, Alaska, Welker Seamount	1050	55.06	-140.41
Paragorgia sp	Paragosp1075756		USNM	1075756	2004	USA, Gulf of Alaska, Alaska, Welker Seamount	1050	55.06	-140.41
Paragorgia sp	Paragosp1075757		USNM	1075757	2004	USA, Gulf of Alaska, Alaska, Pratt Seamount	1093	56.30	-142.47
Paragorgia sp	Paragosp1075759		USNM	1075759	2004	USA, Gulf of Alaska, Alaska, Pratt Seamount	1069	56.17	-142.70
Paragorgia sp	Paragosp1075764		USNM	1075764	2004	USA, Gulf of Alaska, Alaska, Pratt Seamount	920	56.17	-142.70
Paragorgia sp	Paragosp1075769		USNM	1075769	2004	USA, Gulf of Alaska, Alaska, Giacomini Seamount	730	56.42	-146.37
Paragorgia sp	Paragosp1082595		USNM	1082595	2002	USA, Gulf of Alaska, Alaska, Murray Seamount	855	53.89	-148.53
Paragorgia sp	Paragosp1082600		USNM	1082600	2002	USA, Gulf of Alaska, Alaska, Murray Seamount	1376	53.99	-148.50
Paragorgia sp	Paragosp1082644		USNM	1082644	2002	USA, Washington, Warwick Seamount	768	48.05	-132.74
Paragorgia sp	Paragosp1200601		Stone-Sa	1200601	2006	Gulf of Alaska	741	53.98	-162.85
Paragorgia sp	Paragosp121	JX128347	USNM	1122305	2004	British Columbia, Canada		51.20	-130.14
Paragorgia sp	Paragosp1310601		Stone-Sa	1310601	2006	Gulf of Alaska	549	53.73	-163.98
Paragorgia sp	Paragosp159		USNM	1122302	2003	Canada, British Columbia			
Paragorgia sp	Paragosp165		USNM	1122303	2003	Canada, British Columbia, Brooks Penninsula			
Paragorgia sp	Paragosp27294		YPM	27294	2000	Bear Seamount		39.92	-67.44
Paragorgia sp	Paragosp28905		YPM	28905	2003	USA, Atlantic Ocean, Muir Seamount	1713	33.85	-62.66
Paragorgia sp	Paragosp28911		YPM	28911	2003	USA, Atlantic Ocean, Manning Seamount	1550	38.22	-60.46
Paragorgia sp	Paragosp28921		YPM	28921	2003	USA, Atlantic Ocean, Manning Seamount	1579	38.21	-60.53
Paragorgia sp	Paragosp28946		YPM	28946	2003	USA, Atlantic Ocean, Manning Seamount	1597	38.20	-60.53
Paragorgia sp	Paragosp291		USNM	1122306	2004	Canada, British Columbia			
Paragorgia sp	Paragosp34786		YPM	34786	2004	Atlantic Ocean, Manning Seamount, on summit	1336	38.22	-60.51
Paragorgia sp	Paragosp35381		YPM	35381	2004	USA, Manning Seamount, on summit	1337	38.22	-60.51
Paragorgia sp	Paragosp36781		YPM	36781	2004	Bear Seamount, South rim of table top peak	1428-16	39.87	-67.36
Paragorgia sp	Paragosp36910C		YPM	36910C	2004	Bear Seamount, South rim of table top peak	1428-16	39.87	-67.36
Paragorgia sp	Paragosp37148		YPM	37148	2003	USA, Manning Seamount	1483	38.26	-60.55
Paragorgia sp	Paragosp38636		YPM	38636	2004	Bear Seamount, South rim of table top peak	1428-16	39.87	-67.36
Paragorgia sp	Paragosp41849		NIWA	41849	2007	New Zealand, TRIP2494/97	1155.0	-47.0100	175.5800000000
Paragorgia sp	Paragosp44607		NIWA	44607	2007	New Zealand, TRIP2506/24	858.0	-50.0516	174.7150000000
Paragorgia sp	Paragosp54830		USNM	54830	1974	USA, Florida, Straits of Florida, Off Delray Beach	743-761	26.38	-79.60
Paragorgia sp	Paragosp56615		USNM	56615	1974	USA, Florida, Straits of Florida, Off Mile Beach	770-660	25.71	-79.79
Paragorgia sp	Paragosp56690		USNM	56690		USA, Hawaii, Oahu Island, Makapuu Point	366		

Paragorgia sp	Paragospn32830		NIWA	32830	2005	New Zealand, Tangaroa Seamount, P629-4B, KOK0507/4	790	-36.3284'	178.0359955	
Paragorgia sp	ParagospT665A3		USNM	1122225	200?	USA, California, San Juan Seamount	706.6	33.11	-120.96	
Paragorgia sp	Parborea28423		NIWA	28423	1998	New Zealand, 1152/48, Z9275	660	-34.18	162.65	
Paragorgia sp	Parborea28717		YPM	28717	2003	Manning Seamount	1372	38.22	-61.51	
Paragorgia sp	PcfaoteaT668A3		USNM	1122226	200?	USA, California, Little Joe Seamount	2397.8	31.89	-120.05	
Paragorgia sp	PcfaoteaT669A15		USNM	1122227	200?	USA, California, San Marcos Seamount	2061.7	32.64	-121.51	
Paragorgia sp	PcfstepheT665A4		USNM	1122235	200?	USA, California, San Juan Seamount	720.5	33.11	-120.96	
Paragorgia sp	PcfwahineT627A3		USNM	1122238	200?	USA, California, Pioneer Seamount	1771.5	37.40	-123.44	
Paragorgia sp	PcfwahineT630A6		USNM	1122234	200?	USA, California, Rodriguez Seamount	1031.9	34.00	-121.10	
Paragorgia sp	PcfyutliT663A16J4		USNM	1122231	200?	USA, California, Northeast Bank	546.6	32.32	-119.61	
Paragorgia sp		KC788261								
Paragorgia stephencairnsi	2344		RBCM	010-00234-004	2004	British Columbia; VE13978	1194	53.37	-133.31	
Paragorgia stephencairnsi	101010		USNM	1157074	2008	California; Piggy Bank, southern California; DW-026-02	283	33.92	-119.47	
Paragorgia stephencairnsi	1007316		USNM	1007316	2001	British Columbia; Vancouver Island	1168	48.44	-126.38	
Paragorgia stephencairnsi	1124300		USNM	1124300	2006	British Columbia; Vancouver Island, Ohiat Island; OC 06/952	188	48.83	-125.13	
Paragorgia stephencairnsi	Agam		WHOI	Agam	2012	British Columbia; Agamemnon Channel	32	49.72	-124.05	
Paragorgia stephencairnsi	C02		WHOI	C02	2013	British Columbia; Agamemnon Channel; Dive02	41	49.74	-124.03	
Paragorgia stephencairnsi	C03		WHOI	C03	2013	British Columbia; Agamemnon Channel; Dive02	41	49.74	-124.03	
Paragorgia stephencairnsi	C04		WHOI	C04	2013	British Columbia; Agamemnon Channel; Dive02	41	49.74	-124.03	
Paragorgia stephencairnsi	C05		WHOI	C05	2013	British Columbia; Agamemnon Channel; Dive02	41	49.74	-124.03	
Paragorgia stephencairnsi	C100		WHOI	C100	2013	British Columbia; Vancouver Island, Tahsis Inlet; Dive07	40	49.86	-126.67	
Paragorgia stephencairnsi	C101		WHOI	C101	2013	British Columbia; Vancouver Island, Tahsis Inlet; Dive07	40	49.86	-126.67	
Paragorgia stephencairnsi	C102		WHOI	C102	2013	British Columbia; Vancouver Island, Tahsis Inlet; Dive07	40	49.86	-126.67	
Paragorgia stephencairnsi	C104		WHOI	C104	2013	British Columbia; Vancouver Island, Tahsis Inlet; Dive07	40	49.86	-126.67	
Paragorgia stephencairnsi	FOC25		DFO	25	2012	British Columbia; W of Graham Island; 2012-65	204	53.31	-133.03	
Paragorgia stephencairnsi	FOC26		DFO	26	2012	British Columbia; W of Graham Island; 2012-65	221	53.30	-133.04	
Paragorgia stephencairnsi	FOC30		DFO	30	2012	British Columbia; W of Graham Island; 2012-65	318	53.48	-133.07	
Paragorgia stephencairnsi	FOC5		DFO	5	2009	British Columbia; E of Graham Island; 2009-47	201	52.13	-128.90	
Paragorgia stephencairnsi	Paragosp1075741		USNM	1075741	2004	USA, Gulf of Alaska, Alaska, Dickins Seamount	751	54.55	-136.84	
Paragorgia stephencairnsi	Paragosp1092785		USNM	1092785	2005	USA, North Pacific Ocean, Alexander Archipielago, Baranof I:	171	56.19	-135.10	
Paragorgia stephencairnsi	Paragosp200601		Stone-Sa	200601106B01	2006	Gulf of Alaska	479	55.40	-134.83	
Paragorgia stephencairnsi	Paragosp77		USNM	1122304	2004	Canada, British Columbia, Brooks Penninsula		49.94	-128.06	
Paragorgia stephencairnsi	Parborea41106		USNM	1123931	2004	USA, Bering Sea, GOA	417	55.37	-134.78	
Paragorgia stephencairnsi	Parborea411081		USNM	1123930	2004	USA, Bering Sea, GOA	427	54.47	-133.97	
Paragorgia stephencairnsi	ParboreaNew100		USNM	1124301	2006	Canada, British Columbia, Vancouver Island, Barkley Sound	270.22	48.26	-125.01	
Paragorgia stephencairnsi	ParboreaNew545		USNM	1124298	2006	Canada, British Columbia, Vancouver Island, Barkley Sound	309.5	48.15	-125.07	
Paragorgia stephencairnsi			CAS	190438	2010	California, Farallon Escarpement	424	37.74	-123.19	
Paragorgia wahine	Pwahine3326	GQ293314	NIWA	3326	2001	New Zealand, Diabolical seamount, TAN0104/113	900	-42.79	179.99	type
Paragorgia whero	Pwhero3436		NIWA	3436	1998	New Zealand, TRIP1171/12, Z9583	935	-48.03	166.10	type
Paragorgia yutlinux	Pyutlinux1073480	GQ293315	USNM	1073480	2003	Off Vancouver Isl., British Columbia, Canada	846–861	50.23	-128.58	type
Paraminabea aldersladei		JX203767								
Paraminabea aldersladei		KF915662								
Paraminabea aldersladei		KF915665								
Sibogagorgia cauliflora	2036	KC984605	TU-WHC	2036-Oct1	2009	Gulf of Mexico; DC583; J2-454	2440	28.39	-87.39	
Sibogagorgia cauliflora	1122230	GQ293310	USNM	1122230	2006	California; Davidson seamount; dive 945	2502	35.83	-122.61	type

Sibogagorgia cauliflora	Sibogasp1T947A9	GQ293317	USNM	1122229	2006	Davidson Seamount, California, USA	3042.4	35.63	-122.83	
Sibogagorgia cauliflora		GQ293308	USNM	1081143	2004	Derickson Seamount: Alaska: USA	2766	52.98	-161.25	type
Sibogagorgia cauliflora		GQ293309	USNM	54831	1968	Straits of Florida: Havana: Cuba	1638-17	23.55	-82.78	type
Sibogagorgia dennisgordoni		GQ293316	NIWA	3329	1998	New Zealand, 1124/70, Z9228	820	-36.69	176.46	type
Sphaerasclera flammicerebra		JX203765								

Acronyms as follows: National Museum of Natural History, Smithsonian Institution, USA (USNM); The National Institute of Water and Atmospheric Research, New Zealand (NIWA); Museum of Comparative Zoology, Harvard University, USA (MCZ); Muséum National d'Histoire Naturelle, Paris, France (MNHN); Senckenberg Research Institute And Natural History Museum Frankfurt, Germany (SMF); Uppsala University Evolutionsmuseet, Sweden (UUZM); Wakayama Prefectural Museum of Natural History, Japan (WPMNH); Yale Peabody Museum of Natural History, USA (YPM), Department of Fisheries and Oceans Canada (DFO), Woods Hole Oceanographic Institution (WHOI), Temple University (TU), Royal British Columbia Museum (RBCM), California Academy of Scieneces (CAS)

For enabling access to specimens we thank K. Schnabel (NIWA), S. Mills (NIWA), D. Tracey (NIWA), M. Clark (NIWA), A. Rowden (NIWA), S. Cairns (Smithsonian), E. Cordes (Temple U.), A. Quattrini (Temple U.), G. Workman (Department of Fisheries and Oceans Canada - DFO), M. Wyeth (DFO), K. Anderson (DFO), M. Frey (Royal British Columbia Museum - RBCM), H. Gartner (RBCM), L. Watling (U. Hawaii), J. Adkins (CalTech), P. Etnoyer (NOAA), G. Williams (CAS), J. Sanchez (U. Andes), P. Alderslade (CSIRO), A. Andouche (MNHN), A. Andrews (MLML), A. Baco (FSU), A. Baldinger (MCZ), J. A. Boutillier (DFO), S. Davies (DFO), M. Eriksson (UUZM), Y. Imahara (WPMNH), D. Janussen (SMF), E. Lazo-Wasem (YPM), P. Lozouet (MNHN), L. Lundsten (MBARI), B. Stone (NOAA), and many others involved in the planning, collection, and curation.

CHAPTER 6

The genomics of adaptation potential of deep-sea corals to environmental changes

ABSTRACT

Species that live in a wide range of environmental conditions constitute natural experiments of biological adaptations, which can help us to understand possible ecological consequences of environmental changes on ecosystems. A few populations of some deep-sea coral species can be found in shallow (< 45 m) high-latitude fjord environments, where they experience significantly different environmental conditions than their deep relatives. Therefore, these shallow-water populations are believed to inhabit the extremes of the species' physiological tolerances and likely have developed adaptations that enable them to colonize these shallow-water environments. Here, we aim to identify genomic regions that may have enabled the successful adaptation to shallow-water in the deep-sea octocoral species *Paragorgia stephencairnsi*. To characterize the genome-wide genetic diversity of populations of *P. stephencairnsi* found in shallow-water populations and compare it to the genetic diversity from deep-water populations, we performed high-resolution genome-wide scans of single nucleotide polymorphisms. We find patterns of significant population genetic differentiation among the examined populations of *P. stephencairnsi*, which are consistent with the hypothesis that larvae from outer deep populations seeded shallow-water inner fjord populations. Furthermore, we find candidate positive-selection markers shared between parallel comparisons of two shallow populations and a deep populations, and thus identify them as likely candidate makers for genomic regions involved in adaptation to the shallow-water fjord environment. This study lays groundwork for describing the impacts of natural selection on deep-sea coral species in the face of environmental changes.

INTRODUCTION

Species that live in a wide range of environmental conditions constitute natural experiments of biological adaptations, which can help us to understand possible ecological consequences of environmental changes on ecosystems (e.g. Emerson *et al.* 2010). Populations of deep-sea coral species typically live hundreds or

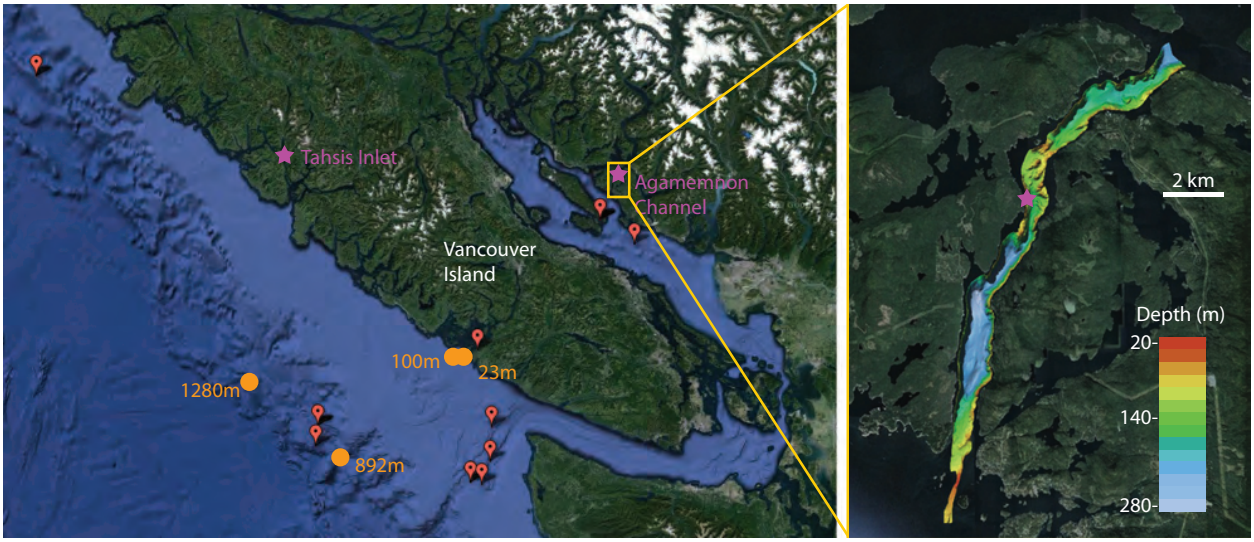
thousands of meters below the surface in the relatively uniform environmental conditions characteristic of the deep-sea (Roberts *et al.* 2009). However, a few populations of some deep-sea coral species can be found in shallow (< 45 m) high-latitude fjord environments, where low light penetration and cool temperatures presumably create suitable living conditions similar to those found in the deep-sea. Despite the seeming similarities between deep-sea and shallow fjord environments there are significant differences that make these shallow fjords a novel environment for deep-sea species, namely lower hydrostatic pressure and significantly greater ranges of variability for temperature, pH, salinity, current speeds, and sedimentation rates. Therefore, these shallow populations of predominantly deep-sea coral species are believed to inhabit the extremes of the species' physiological tolerances.

Differences in environmental conditions over space and time can have strong selective effects on natural populations by modifying the survival and reproductive success of individuals, and thus altering genetic composition of the populations and their ability to respond to environmental changes (e.g., Prada & Hellberg 2013). When isolated populations of a species are exposed to similar selective pressures, e.g. deep-sea coral populations in semi-enclosed shallow environments such as inlets and fjords, they tend to develop similar solutions to common challenges – a process known as parallel adaptation (e.g., Chan *et al.* 2010; Hohenlohe *et al.* 2010; Jones *et al.* 2012; Miller *et al.* 2012). Hohenlohe *et al.* (2010) present an exemplary case exploring the genome-wide consequences of this evolutionary process. In that study the authors investigate the parallel adaptation to freshwater environments in marine stickleback populations by performing high-resolution genomic scans of single nucleotide polymorphisms (SNPs) from ancestral marine and derived freshwater populations. When genome-mapped data from marine populations are compared to freshwater populations it is possible to detect specific regions in the genome that had more differentiation than what is expected under neutrality. Common regions of differentiation across multiple populations are identified as strong candidates for parallel adaptation of populations of a marine species to freshwater. Hohenlohe *et al.* (2010) show that several genomic regions identified through this method co-localize with previously identified quantitative trait loci (regions of the genome that account for particular observable characteristics of organisms), thus demonstrating the usefulness of this approach for the identification of ecologically important genes.

Bubblegum corals (Paragorgiidae, Octocorallia) are among the most abundant and widely distributed benthic foundation species in deep-water ecosystems worldwide (Roberts *et al.* 2009; Wating *et al.* 2011). They play an important ecological role, akin to the structural role of large trees in a rainforest, by generating three-dimensional habitats for a great number of micro- and macro-organisms (Buhl-Mortensen & Mortensen 2004; Auster *et al.* 2005; Buhl-Mortensen & Mortensen 2005; DeVogelaere *et*

al. 2005; Nedashkovskaya *et al.* 2005). Populations of the bubblegum coral species *Paragorgia stephencairnsi* Sánchez, 2005 (*sensu* Herrera and Shank (Chapter 5)) are typically found in the ocean at depths greater than 200 meters, attached to hard grounds on the continental shelf and slope, and seamounts along the western coast of North America (Sánchez 2005). A few populations of this species inhabit shallow fjords (as shallow as 30 meters) at the northern boundary of its distribution. Recent phylogenomic evidence shows *P. stephencairnsi* evolved from deep-sea ancestors (Herrera & Shank Chapter 5). Thus, the most parsimonious scenario is that shallow-water populations of *P. stephencairnsi* in high-latitude fjords originated from colonization seeded by deeper populations.

a)



b)

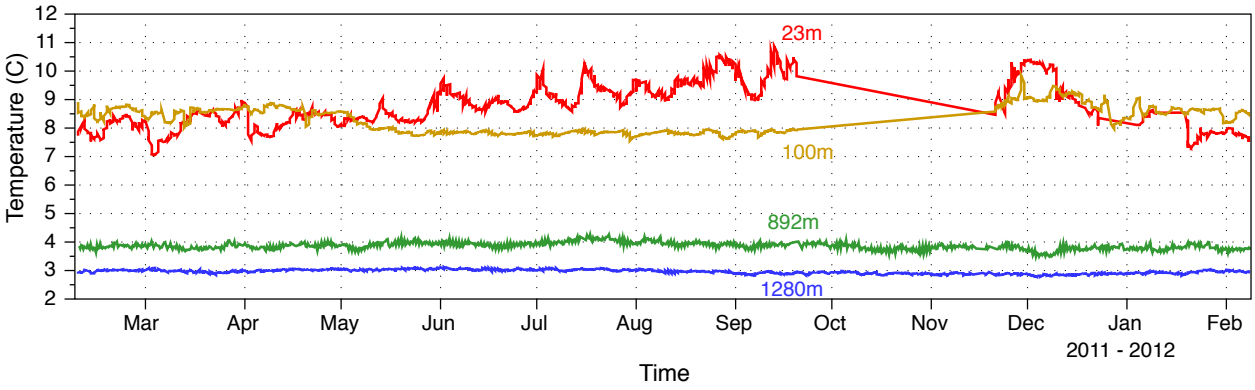


Figure 1. a) Geographic location of *P. stephencairnsi* populations off British Columbia, Canada. **Left** map shows sites where samples have been obtained. Purple stars indicate locations of shallow populations (accesible via scuba diving). Red flags indicate the locations of deep populations. Orange dots indicate observatory nodes from the NorthEast Pacific Time-Series Undersea Networked Experiments (NEPTUNE) monitoring environmental variables

at 23m, 100m, 892m, and 1280m depths. **Right** map shows the bathymetry of the glacially formed Agamemnon Channel, one of the primary shallow-water fjord collection sites. **b)** Temperature records from monitoring nodes at four different depths off the North Western coast of North America. Location of nodes shown in Fig. 3a. Data courtesy of NEPTUNE ocean network observatory (Canada).

In general, fjords –long and narrow coastal sea inlets formed by glaciers – can be thought of as semi-enclosed marine basins as they are limited at their sides by landmasses. The seafloor in fjords dramatically shoals at their mouths because of the paleoglaciers' terminal moraines (see Agamemnon Channel example in Fig. 1a). The geological characteristics of fjords constrain the circulation and exchange between these semi-enclosed environments and the 'outer' ocean environment. Thus, the amounts of gene flow between 'outer' deep and 'inner' shallow populations of marine organisms, such as *Paragorgia stephencairnsi*, are likely limited by the these same geomorphic features.

The main goal of this study is to identify and characterize potential genomic regions that have enabled the successful colonization of shallow-water environments from the deep-sea by *P. stephencairnsi*. We hypothesize that semi-isolated shallow water populations of *P. stephencairnsi* in different fjord systems have independently evolved adaptations in parallel to cope with distinct highly variable conditions of the surface ocean. As a result the expectation is that, in each population, the environment has independently selected common sets of genetic diversity. This would leave characteristic signatures of parallel differentiation in their genomes. Here, we perform high-resolution genome-wide scans of SNPs to characterize the genome-wide genetic diversity of populations of *P. stephencairnsi* found in shallow-water populations, and compare it to the genetic diversity from deep-water populations. Through these comparisons we identify patterns of differentiation that would be indicative of non-random evolutionary processes of natural selection and adaptation and lay groundwork for describing the impacts of natural selection on deep-sea coral species in the face of environmental changes.

METHODS

Specimens of *P. stephencairnsi* from shallow water populations (less than 45m; Agamemnon Channel and Tahsis Strait) were collected in May 4-17, 2013 in British Columbia, Canada, during 8 decompression SCUBA dives. Specimens from deep populations around Vancouver Island were collected in various oceanographic expeditions by collaborators at Memorial University, the Department of Fisheries and Oceans Canada, the Royal British Columbia Museum, and the National Oceanographic and Atmospheric Administration (Table 1).

Table 1. Collection information for the specimens used in this study.

ID	Collection	Catalog Number	Date	Locality	Depth (m)	Lat.	Lon.	Population
Agam	WHOI	Agam	2012	Agamemnon Channel	32	49.72	-124.05	Agamemnon
C02	WHOI	C02	2013	Agamemnon Channel; Dive02	41	49.74	-124.03	Agamemnon
C03	WHOI	C03	2013	Agamemnon Channel; Dive02	41	49.74	-124.03	Agamemnon
C04	WHOI	C04	2013	Agamemnon Channel; Dive02	41	49.74	-124.03	Agamemnon
C05	WHOI	C05	2013	Agamemnon Channel; Dive02	41	49.74	-124.03	Agamemnon
C07	WHOI	C07	2013	Agamemnon Channel; Dive02	41	49.74	-124.03	Agamemnon
C08	WHOI	C08	2013	Agamemnon Channel; Dive02	41	49.74	-124.03	Agamemnon
C11	WHOI	C11	2013	Agamemnon Channel; Dive02	41	49.74	-124.03	Agamemnon
C12	WHOI	C12	2013	Agamemnon Channel; Dive02	41	49.74	-124.03	Agamemnon
C100	WHOI	C100	2013	Vancouver Island, Tahsis Inlet; Dive07	40	49.86	-126.67	Tahsis
C101	WHOI	C101	2013	Vancouver Island, Tahsis Inlet; Dive07	40	49.86	-126.67	Tahsis
C102	WHOI	C102	2013	Vancouver Island, Tahsis Inlet; Dive07	40	49.86	-126.67	Tahsis
C103	WHOI	C103	2013	Vancouver Island, Tahsis Inlet; Dive07	40	49.86	-126.67	Tahsis
C104	WHOI	C104	2013	Vancouver Island, Tahsis Inlet; Dive07	40	49.86	-126.67	Tahsis
C105	WHOI	C105	2013	Vancouver Island, Tahsis Inlet; Dive07	40	49.86	-126.67	Tahsis
C106	WHOI	C106	2013	Vancouver Island, Tahsis Inlet; Dive07	40	49.86	-126.67	Tahsis
C107	WHOI	C107	2013	Vancouver Island, Tahsis Inlet; Dive07	40	49.86	-126.67	Tahsis
FOC25	DFO	25	2012	W of Graham Island; 2012-65	204	53.31	-133.03	Deep
FOC26	DFO	26	2012	W of Graham Island; 2012-65	221	53.30	-133.04	Deep
FOC30	DFO	30	2012	W of Graham Island; 2012-65	318	53.48	-133.07	Deep
FOC5	DFO	5	2009	E of Graham Island; 2009-47	201	52.13	-128.90	Deep
L139	MU	R1513-L1-0039		S of Texada Island	268	49.50	-124.17	Deep
L219	MU	R1513-L2-0019		S of Texada Island	267	49.50	-124.17	Deep
L341	MU	R1513-L3-0041		S of Texada Island	268	49.50	-124.17	Deep
PR27	MU	R1513-PR-0027		S of Texada Island	270	49.50	-124.17	Deep
1122479	USNM	1122479	2008	LaPush, west of, Washington	269	48.13	-125.10	Deep
1124300	USNM	1124300	2006	Vancouver Island, Ohiat Island	188	48.83	-125.13	Deep

Acronyms as follows: National Museum of Natural History, Smithsonian Institution, USA (USNM); Department of Fisheries and Oceans Canada (DFO), Woods Hole Oceanographic Institution (WHOI), Memorial University (MU).

Molecular laboratory methods

To characterize the genome-wide genetic diversity of populations of *P. stephencairnsi,* we performed high-resolution genomic scans and identified single nucleotide polymorphisms (SNPs) from restriction site-associated DNA markers (RAD tags) (Baird *et al.* 2008; Hohenlohe *et al.* 2010). In short, the RAD sequencing method consists of: 1) the digestion of genomic DNA for each individual with a restriction enzyme; 2) ligation of the resulting fragments to sequencing adapters with unique barcodes for each individual; 3) size-selection and enrichment of the fragments successfully ligated to the adapters; and 4) sequencing via a high-throughput platform (Illumina HiSeq 2000). We performed RAD sequencing with the 6-cutter restriction enzyme PstI, which is predicted to cut between 32,000 and 110,000 times in the

genome of an octocoral (Herrera & Shank Chapter 5). This predicted range was obtained using the observed frequency of the PstI recognition sequence, and its probability using a trinucleotide composition model, in the genomes of the cnidarians *Nematostella vectensis, Acropora digitifera, Hydra vulgaris,* and *Alatina moseri* (Herrera *et al.* Chapter 2). Genome size range of 0.3-0.5 pg was used based on observations obtained through flow cytometry in gorgoniid octocorals by Luisa Dueñas at the Universidad de los Andes, Bogotá, Colombia (personal communication).

Total genomic DNA was purified from specimens as in Herrera and Shank (Chapter 5) by: (1) digesting the tissue in 2% CTAB buffer (Teknova) with proteinase K and RNAse A/T1 (Fermentas) for 1 hour, (2) separating nucleic acids with chloroform: isoamyl alcohol (24:1) (Fermentas) and phenol: chloroform: isoamyl alcohol (25:24:1, Tris buffered at pH 8.0) (Fermentas), (3) precipitating nucleic acids with 100% ethanol (1:1 volume ratio), and (4) washing the precipitate twice with 70% ethanol. Concentration-normalized genomic DNA was submitted to Floragenex Inc. (Eugene, OR) for library preparation and RAD sequencing. Libraries were sequenced by 48-multiplex, using 10-base pairs long barcodes, on a single lane of an Illumina Hi-Seq 2000 sequencer.

Data filtering

Sequence reads were de-multiplexed and quality filtered with the *process_radtags* program from the package Stacks v1.19 (Catchen *et al.* 2013b). Barcodes and Illumina adapters were excluded from each read and length was truncated to 91bp (-t 91) Reads with ambiguous bases were discarded (-c). Reads with an average quality score below 10 (-s 10) within a sliding window of 15% of the read length (-w 0.15) were discarded (-r). The rescue barcodes and RAD-tags algorithm was enabled (-r).

De novo loci assembly

We performed *de novo* assemblies of RAD loci using the *denovo_map* pipeline in Stacks. A minimum depth of three reads per stack was enforced (-m 3). Significantly high-repetitive stacks were discarded by implementing the deleveraging algorithm (-t), as these likely represent sequencing errors, duplications or repetitive regions. No mismatches among loci were allowed when creating the catalog of all the loci identified among the sampled individuals (-n 0). A maximum number of two mismatches was allowed among loci within each individual (-M 2). The maximum number of stacks at a single locus was set to three (--max_locus_stacks 3).

Demographic Inferences

We estimated population genetic descriptive statistics per SNP (nucleotide diversity π, proportion of polymorphic loci, observed heterozygosity, minor allele frequency, number of private alleles, inbreeding index $F_{IS,}$ and population differentiation index F_{ST}) using the program *populations* of Stacks. We only analyzed loci that were present in all populations of each species (-p) and in all individuals in each population (-r). We calculated population F_{ST} values utilizing a *p_value* filter (-f) to keep only significant estimates (α=0.05). We exported SNP data in *genpop* format, keeping only one SNP per locus to avoid violating the assumption of independence among markers in downstream analyses.

To summarize the variation in the SNP data among individuals and populations we a performed principal component analyses (PCA) as in (Reitzel *et al.* 2013), using the program *smartpca* from the package Eigensoft v5.0 (Patterson *et al.* 2006; Price *et al.* 2006). We evaluated the significance of the identified principal components through Tracy-Widom statistics (Tracy & Widom 1994; Johnstone 2001) and evaluated the statistical significance of the differences between populations with a chi-square test.

Candidate adaptation markers

We identified markers linked to candidate genomic regions involved in adaptation to shallow water environments from the deep-sea by detecting F_{ST} outliers (i.e., allelic frequencies between deep and shallow populations that show greater differentiation than expected under a neutral model of evolution, characterized by the accumulation of random mutations that do not affect survival and reproduction of organisms) with the program LOSITAN (Beaumont & Nichols 1996; Antao *et al.* 2008). Population genomics theory predicts that these outlier variants will be indicative of genomic regions containing genes or regulatory elements that have been subject to natural selection (Lewontin & Krakauer 1973; Beaumont & Balding 2004). SNP positions with outlier F_{ST} values (those above the 97.5 percentile of the neutral distribution of F_{ST}) were considered indicative of loci subject to natural selection. We considered outliers shared among shallow populations as indicative of parallel adaptations to the shallow-water environment, whereas outliers unique to a particular population were considered as likely indicators of local adaptations (see Hohenlohe *et al.* 2010).

Genomic sequences of markers under potential positive selection were scanned for functionality by querying against annotated databases of gene models from cnidarian genomes (*Nematostella vectensis* and *Hydra magnipapillata*) using BLAST searches at the U.S. National Center for Biotechnology Information (NCBI) databases. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes

(KEGG) pathway assignments were attempted for each gene-model match using the program Blast2GO (Conesa *et al.* 2005).

Table 2. RAD sequencing results, filtering and *de novo* assebly statistics.

Population	RAD-seq data file ID	Total sequenced reads	Retained reads	Percentage of retained reads after filtering	Number of Stacks	Mean coverage depth	S.D. of coverage depth
Agamemnon	PAR_Agam_BC	4,322,564	4,062,469	94.0	142,874	25.3	151.6
Agamemnon	PAR_C02_BCS	2,149,938	1,594,818	74.2	110,403	10.8	35.7
Agamemnon	PAR_C03_BCS	2,815,327	2,264,951	80.5	126,442	13.7	73.8
Agamemnon	PAR_C04_BCS	3,240,713	2,754,373	85.0	133,631	15.9	87.3
Agamemnon	PAR_C05_BCS	3,465,396	2,972,548	85.8	136,642	16.9	104.3
Agamemnon	PAR_C07_BCS	2,827,394	2,446,736	86.5	129,688	14.5	59.6
Agamemnon	PAR_C08_BCS	2,552,442	2,104,566	82.5	125,331	12.5	50.9
Agamemnon	PAR_C11_BCS	2,763,403	2,286,010	82.7	128,412	13.6	58.6
Agamemnon	PAR_C12_BCS	2,390,479	2,062,541	86.3	124,553	12.5	53.5
Deep	PAR_1122479_BC	4,607,551	4,331,474	94.0	142,445	27.2	108.6
Deep	PAR_1124300_WA	3,843,250	3,627,391	94.4	136,241	23.7	152.2
Deep	PAR_FOC25_BCD	5,479,109	4,693,804	85.7	152,378	24.2	102.9
Deep	PAR_FOC26_BCD	5,242,949	4,528,629	86.4	210,583	14.9	76.2
Deep	PAR_FOC30_BCD	4,417,520	3,781,973	85.6	174,618	15.8	116.6
Deep	PAR_FOC5_BCD	2,834,377	2,455,041	86.6	135,156	13.3	129.2
Deep	PAR_L139_BCD	2,928,518	2,392,739	81.7	129,058	14.1	62.6
Deep	PAR_L219_BCD	1,752,674	1,289,295	73.6	96,904	9.5	51.0
Deep	PAR_L341_BCD	3,253,979	2,761,200	84.9	133,647	16.1	75.8
Deep	PAR_PR27_BCD	3,536,670	2,942,112	83.2	135,368	17.0	72.3
Tahsis	PAR_C100_BCS	5,998,914	4,984,766	83.1	149,979	26.1	90.1
Tahsis	PAR_C101_BCS	5,332,619	4,652,535	87.2	150,525	24.2	142.6
Tahsis	PAR_C102_BCS	4,347,757	3,710,381	85.3	141,586	20.5	73.7
Tahsis	PAR_C103_BCS	2,272,200	1,706,852	75.1	114,196	11.0	38.5
Tahsis	PAR_C104_BCS	2,720,994	2,203,568	81.0	124,966	13.5	56.5
Tahsis	PAR_C105_BCS	2,778,573	2,003,812	72.1	123,158	12.1	45.9
Tahsis	PAR_C106_BCS	2,565,150	2,116,517	82.5	123,271	13.1	59.1
Tahsis	PAR_C107_BCS	2,697,054	2,189,475	81.2	125,420	13.3	45.6
	AVERAGE	3,449,538	2,922,984	83.7	135,462	16.5	80.5
	S.D.	1,136,671	1,066,781	5.6	21,081	5.2	34.5

RESULTS

RAD-seq produced high-quality sequence data

We generated restriction site associated DNA sequence (RAD-seq) data for 27 individuals of *P. stephencairnsi* collected in the British Columbia region (Table 2). We obtained approximately 3.5±1.1

(mean ± standard deviation) million sequence reads per individual (100bp length), with individual values ranging from 1.75 to 6.0 million reads. Approximately 83.7±5.6% of these were retained after quality filters. *De novo* loci assemblies produced approximately 135±21 thousand unique sequence stacks per individual, with a mean coverage depth of 16.5±5.2X.

Table 3. Population genetic statistics calculated from only variant positions, and from both variant and fixed positions. Values indicate means ± standard deviation

Variant positions

Population	Private alleles	Variant sites	% polym. sites	Major allele frequency	Observed heterozygosity	Nucleotide diversity (π)	F_{IS}
Agamemnon	2,713	17,074	65.23	0.8737±0.1517	0.2149±0.2604	0.1858±0.1884	-0.0612±0.2404
Tahsis	2,000	17,071	55.54	0.8793±0.1578	0.2000±0.2737	0.1743±0.1975	-0.0468±0.2717
Deep	3,046	17,062	69.35	0.8697±0.1523	0.2102±0.2565	0.1904±0.1865	-0.0340±0.2851

All positions (variant and fixed)

Population	Private alleles	Sites	% polym. sites	Major allele frequency	Observed heterozygosity	Nucleotide diversity (π)	F_{IS}
Agamemnon	2,713	1,873,995	0.59	0.9988±0.0200	0.0020±0.0316	0.0017±0.0245	-0.0006±0.0245
Tahsis	2,000	1,873,999	0.51	0.9989±0.0200	0.0018±0.0316	0.0016±0.0245	-0.0004±0.0265
Deep	3,046	1,873,984	0.63	0.9988±0.0200	0.0019±0.0316	0.0017±0.0245	-0.0003±0.0283

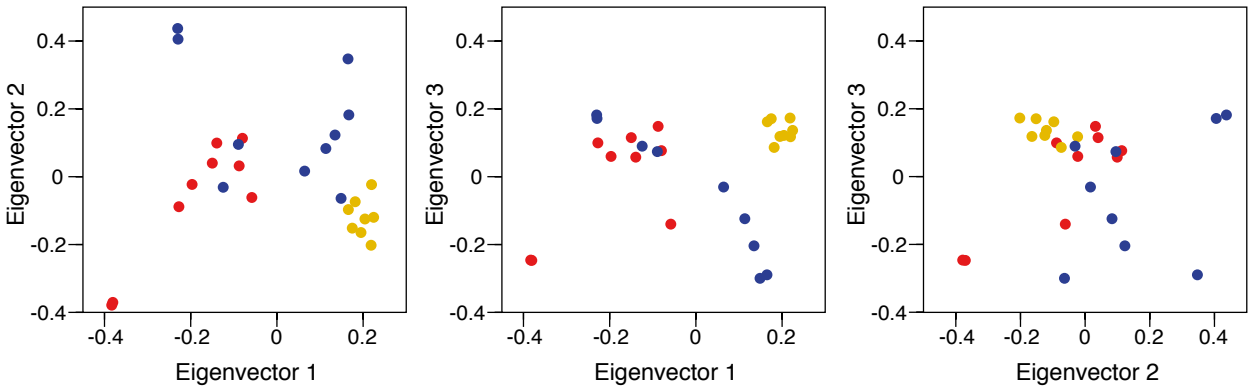


Figure 2. Principal component analysis of genetic variation from SNP in *Paragorgia stephencairnsi* from British Columbia. Each dot represents an individual. Colors indicate the source population: Agamemnon (red), Tahsis (yellow), and Deep (blue). The three principal axes of variation (eigenvectors) are shown.

Significant population differentiation among all populations

There were 10,920 loci shared among all individuals in all populations. These loci contained over 17 thousand SNPs, for an average of 1.56 SNP per locus (Table 3). The three largest axes of variation (eigenvectors) identified from the principal components analysis of SNP data revealed clear separation between the two shallow-water populations Tahsis and Agamemnon, but evident overlap between the individuals from the Deep population and both of the shallow (Fig. 3). All eigenvectors were statistically significant ($P<0.001$, $P=0.002$, and $P=0.031$ for eigenvectors 1, 2 and 3, respectively; $\alpha=0.05$). All differences among populations were also statistically significant (Agamemnon vs. Tahsis: $\chi^2=40.1$, $P<0.001$; Agamemnon vs. Deep: $\chi^2=16.4$, $P=0.003$; Tahsis vs. Deep: $\chi^2=30.3$, $P<0.001$; $\alpha=0.05$). Mean pairwise F_{ST} values indicate that population differentiation is significantly greater between the two shallow-water populations Tahsis and Agamemnon ($F_{ST}=0.0519$, $P<0.05$), than between either shallow-water population and the Deep population (Agamemnon vs. Deep: $F_{ST}=0.0354$, $P<0.05$; Tahsis vs. Deep: $F_{ST}=0.0430$, $P<0.05$).

Differences in nucleotide diversity between shallow and deep-populations

Summary population genetic statistics (private alleles, percentage of polymorphic sites, and nucleotide diversity) calculated from variant positions revealed higher diversity in the Deep population than in either of the shallow-water populations. These same metrics indicated that the Tahsis shallow-water population has the lowest genetic diversity of all three examined populations. Differences in population genetic diversity were not evident when summary population genetic statistics were calculated from both fixed and variant positions.

Differences in allele frequency distributions between shallow and deep populations

Minor allele frequency spectra show that a majority of the alleles in each population have low frequencies (Fig. 2), as expected for population near mutation-drift equilibrium. Shallow-water populations show a small modal shift towards higher allelic frequencies (particularly in Tahsis) and a noticeable increase of intermediate frequency alleles, compared to the deep population. F_{IS} distributions for all populations were centered on zero, with a tendency to negative values, indicating random mating in populations and a slight excess of heterozygotes.

Candidate adaptation markers

Outlier analyses between shallow-water populations and the Deep population found 733 SNPs candidate positive-selection markers when comparing Agamemnon vs. Deep, and 261 when comparing Tahsis vs.

Deep. Of these, 63 SNPs candidate positive selection markers were shared in both comparisons, and thus are considered candidate makers for adaptive genomic regions.

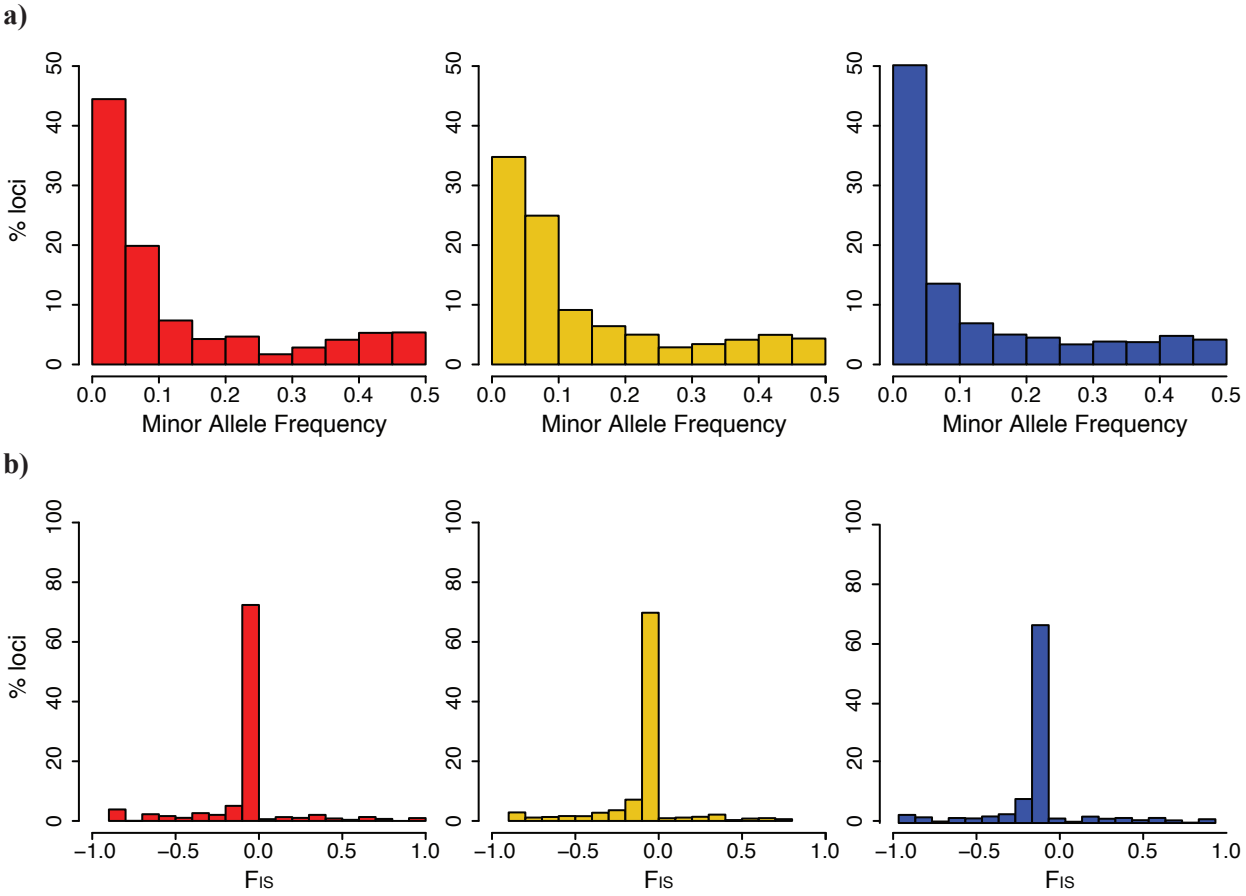


Figure 3. Frequency distributions of population genetic summary statistics for populations of *Paragorgia stephencairnsi* from British Columbia. **a)** Minor allele frequency spectra for SNPs loci. **b)** Frequency spectra of Wright's inbreeding index F_{IS} values calculated for each SNP locus. Colors indicate the source population: Agamemnon (red), Tahsis (yellow), and Deep (blue).

From the markers identified as potentially adaptive, 16 produced Blast matches to the non-redundant nucleotide NCBI database (Table 4). None produced matches to available protein or gene ontology databases. Many of the matches corresponded to coding DNA sequences of unknown function and uncharacterized mRNAs. The top three marker matches, in terms of Blast E-values (the number of expected false-positive matches in a database of given size), correspond to mRNAs of *Hydra magnipapillata:* a mitogen-activated protein kinase 7-like mRNA (RAD locus catalog 17484) potentially involved in cellular cycles and cell differentiation; a phosphoribosylformylglycinamidine synthase partial mRNA (RAD locus catalog 23843) potentially involved in purine metabolism; and a zinc finger protein

41-like mRNA (RAD locus catalog 83962) potentially a transcription factor associated with meiosis in spermatogenesis.

Table 4. Blast results for the candidate positive selection markers shared between comparisons of shallow-water populations vs. Deep population. Only results that produced hits in the non-redundant nucleotide NCBI database are shown.

RAD locus catalog name	Blast hit species	Blast hit sequence description	Blast E-value	Blast hit accession	Blast hit % similarity
17484	*Hydra magnipapillata*	mitogen-activated protein kinase 7-like partial mrna	1.00E-08	XM004210406	75
23843	*Hydra magnipapillata*	phosphoribosylformylglycinamidine synthase partial mrna	3.60E-08	XM001641610	77
83962	*Hydra magnipapillata*	zinc finger protein 41-like mrna	2.80E-03	XM002162877	80
57130	*Hydra magnipapillata*	uncharacterized loc100207636 partial mrna	9.60E-03	XM002158158	79
26651	*Nematostella vectensis*	protein partial mrna	9.60E-03	XM001627193	83
843	*Nematostella vectensis*	gene for complete cds	1.20E-01	BR000671	79
36131	*Nematostella vectensis*	protein partial mrna	4.10E-01	XM001629268	81
86101	*Hydra magnipapillata*	uncharacterized loc101234456 mrna	1.40E+00	XM004211848	78
101291	*Hydra magnipapillata*	uncharacterized loc101240885 mrna	1.40E+00	XM004206580	80
22838	*Hydra magnipapillata*	lysine-specific demethylase 6a-like mrna	1.40E+00	XM002167270	83
67285	*Hydra magnipapillata*	uncharacterized loc100199733 mrna	1.40E+00	XM002159843	80
76378	*Hydra magnipapillata*	uncharacterized loc100202739 mrna	5.00E+00	XM002162521	77
103847	*Hydra magnipapillata*	uncharacterized loc100207904 mrna	5.00E+00	XM002160483	75
71485	*Hydra magnipapillata*	uncharacterized loc100209560 mrna	5.00E+00	XM002160476	78
8691	*Nematostella vectensis*	protein partial mrna	5.00E+00	XM001627130	80
27970	*Nematostella vectensis*	protein complete cds	5.00E+00	XM001639122	79

DISCUSSION

Fjord environments effectively isolate marine populations

In this study, we find significant population genetic differentiation among all examined populations of *Paragorgia stephencairnsi*. This genetic differentiation is particularly marked between the shallow-water fjord populations as indicated by the lack of overlap among groups in the principal components analysis (PCA) and the elevated population differentiation F_{ST} value. Our results indicate that fjords act as semi-enclosed basins effectively isolating coral populations living within them. The evidence we present conforms to previous results from population genetic studies in pelagic and benthic organisms such as calanoid copepods (Bucklin *et al.* 2000), glacier lanternfish (Suneetha & Salvanes 2001), sea stars (Perrin *et al.* 2004), Pacific cod (Cunningham *et al.* 2009), sprat (Glover *et al.* 2011), and Pacific herring (Wildes

et al. 2011), which show significant population genetic structuring in small spatial scales among fjord populations, and with respect to open ocean populations.

Shallow-water fjord populations originated from the deep

Taken together, our results are consistent with the hypothesis that larvae from outer deep populations seeded shallow-water inner fjord populations. Smaller genetic diversity found in the shallow-water fjord populations relative to those from the deep population, indicated by smaller spread in the PCA and smaller nucleotide diversity values (π), is suggestive of larger effective population size (N_e) in the deep population (given the relationship $\pi=4N_e\mu$, assuming equal mutation rates μ and mutation-drift equilibrium). The evident, although small, overlap in the PCA between the large deep population and each one of the smaller shallow-water fjord populations suggests limited gene flow from the deep. This colonization from the deep must have occurred relatively recently, given that present-day North American fjords were predominantly occupied by massive glaciers or exposed to the atmosphere due to sea-level change during the last glacial maximum 33,000-14,500 years ago (Clague & James 2002; Clark & Mix 2002; Clark *et al.* 2009). However, we suggest enough time may have passed since colonization because allele frequency distributions are not strongly skewed towards fixation, as expected for young non-equilibrium populations founded by few individuals (e.g., Catchen *et al.* 2013a). The dynamics of population structuring between shallow-water inner fjord and outer deep populations of *P. stephencairnsi* over time are unknown. However, compelling evidence from other species suggests that these patterns of differentiation between fjord and open water populations can be stable over at least hundreds of years (Harnstrom *et al.* 2011).

Natural selection in shallow-water fjord environments

The differences in conditions between open water and shallow fjord environments (e.g. Fig. 1b) can also act as barriers for gene flow, further limiting the amount of gene flow between shallow and deep-sea populations. Compared to deep populations, the shallow-water fjord populations are exposed to lower hydrostatic pressure and significantly greater variability ranges of temperature, pH, salinity, current speeds, and sedimentation rates influenced by marked seasonality (observations from the NEPTUNE and VENUS time series). These conditions may not only act as barriers for gene flow but also as selective forces. The observed F_{IS} distributions with a tendency to negative values and the higher percentage of loci of intermediate allelic frequencies in shallow-water populations, relative to the deep population, may be the result of natural selection (e.g. heterozygote advantage or balancing selection). Alternatively, demographic processes, such as bottlenecks, may also produce similar patterns (Luikart *et al.* 1998).

The comparative approach between shallow-water populations and the deep ancestral population of *P. stephencairnsi*, allowed us to identify, for the first time, potential markers of parallel adaptation to the shallow-water environment in a deep-sea organism. Although only three markers were mapped to known cnidarian functional regions, there is the potential to identify more genomic regions and link them to potential functions and mechanisms as more genomic resources become available (Hohenlohe *et al.* 2010; Reitzel *et al.* 2013). Significant environmental changes are occurring due to anthropogenic CO_2 emissions (Hoegh-Guldberg & Bruno 2010). Despite efforts to understand the effect of these changes on marine species, little is known about the adaptive mechanisms that would allow them to survive over ecological and evolutionary time scales. Shallow-water populations of deep-sea coral species have already adapted to deal with the environmental extremes of the surface ocean, thus they could constitute pre-adapted populations that could expand their range to deeper water in the case of significant environmental changes at depth, thereby seeding future deep-sea ecosystems. On the other hand, shallow-water populations of deep-sea corals may live at the tolerance boundary of their species, and thus could face habitat shifts and local extinction in the near future. A better understanding of the adaptive potential of these corals will allow us to assess the possible impacts of climate change on the diverse but vulnerable ecosystems supported by these habitat-forming corals.

Future work

To increase the confidence in our demographic and natural selection inferences we plan to perform RAD sequencing on additional available individuals from deep and shallow populations. Additionally, the potential markers of parallel adaptation to the shallow-water environment will be mapped to the draft genome sequence of an *P. stephencairnsi* individual, which will be generated using high-throughput sequencing and routine algorithms developed to assemble full genomes from short sequence-reads (following Gnerre *et al.* 2011). High-resolution genomic scans generated by RAD-seq provide genotypic data for tens of thousands of SNPs, thus allowing the creation of genome-wide distributions of F_{ST} and other population genetic summary statistics, which allow further identification of candidate potential genomic regions and elements involved in parallel adaptation to the shallow-water environment (see Hohenlohe *et al.* 2010).

ACKNOWLEDGEMENTS

This research was supported by the National Geographic Society/Waitt Foundation (W285-13 to SH); the National Oceanic and Atmospheric Administration (NA09OAR4320129 to TS); the National Science Foundation (OCE-1131620 to TS); the National Aeronautics and Space Administration (NNX09AB76G

to TS); and the Academic Programs Office (Ocean Ventures Fund to SH), the Ocean Exploration Institute (Fellowship support to TMS) and the Ocean Life Institute of the Woods Hole Oceanographic Institution (WHOI). For enabling access to key specimens we thank S. Cairns (Smithsonian), G. Workman (Department of Fisheries and Oceans Canada - DFO), M. Wyeth (DFO), K. Anderson (DFO), M. Frey (Royal British Columbia Museum - RBCM), H. Gartner (RBCM), Ed Bowlby (NOAA, Olympic Coast National Marine Sanctuary), E. Edinger (Memorial U.) and J. Sanchez (U. Andes). We also thank E. O'Brien (WHOI), D. Forsman (WHOI), J. Fellows (WHOI), J. & S. Schooner, K. Heylar, and N. McDaniel for invaluable assistance during scuba diving fieldwork in British Columbia (DFO scientific license FIN130270). We also thank A. Tarrant, A. Reitzel, and E. Bors for providing helpful comments that improved this manuscript.

REFERENCES

Antao T, Lopes A, Lopes R, Beja-Pereira A, Luikart G (2008) LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics* **9**, 323.

Auster PJ, Moore J, Heinonen KB, Watling L (2005) A habitat classifi cation scheme for seamount landscapes: assessing the functional role of deep-water corals as fish habitat. In: *Cold-water Corals and Ecosystems* (eds. Freiwald AR, Roberts JM), pp. 761-769. Springer-Verlag, Berlin Heidelberg.

Baird NA, Etter PD, Atwood TS*, et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**, e3376.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**, 969-980.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London Series B-Biological Sciences* **263**, 1619-1626.

Bucklin A, Kaartvedt S, Guarnieri M, Goswami U (2000) Population genetics of drifting (*Calanus* spp.) and resident (*Acartia clausi*) plankton in Norwegian fjords. *Journal of Plankton Research* **22**, 1237-1251.

Buhl-Mortensen L, Mortensen PB (2004) Crustaceans associated with the deep-water gorgonian corals *Paragorgia arborea* (L., 1758) and *Primnoa resedaeformis* (Gunn., 1763). *Journal of Natural History* **38**, 1233-1247.

Buhl-Mortensen L, Mortensen PB (2005) Distribution and diversity of species associated with deep-sea gorgonian corals off Atlantic Canada. In: *Cold-Water Corals and Ecosystems* (eds. Freiwald AR, Roberts JM), pp. 849-879. Springer-Verlag, Berlin Heidelberg.

Catchen J, Bassham S, Wilson T*, et al.* (2013a) The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. *Molecular Ecology* **22**, 2864-2883.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013b) Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**, 3124-3140.

Chan YF, Marks ME, Jones FC*, et al.* (2010) Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* **327**, 302-305.

Clague JJ, James TS (2002) History and isostatic effects of the last ice sheet in southern British Columbia. *Quaternary Science Reviews* **21**, 71-87.

Clark PU, Dyke AS, Shakun JD*, et al.* (2009) The Last Glacial Maximum. *Science* **325**, 710-714.

Clark PU, Mix AC (2002) Ice sheets and sea level of the Last Glacial Maximum. *Quaternary Science Reviews* **21**, 1-7.

Conesa A, Gotz S, Garcia-Gomez JM*, et al.* (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674-3676.

Cunningham KM, Canino MF, Spies IB, Hauser L (2009) Genetic isolation by distance and localized fjord population structure in Pacific cod (*Gadus macrocephalus*): limited effective dispersal in the northeastern Pacific Ocean. *Canadian Journal of Fisheries and Aquatic Sciences* **66**, 153-166.

DeVogelaere AP, Burton EJ, Trejo T*, et al.* (2005) Deep-sea corals and resource protection at the Davidson Seamount, California, U.S.A. In: *Cold-water Corals and Ecosystems* (eds. Freiwald AR, Roberts JM), pp. 1189-1198. Springer-Verlag, Berlin Heidelberg.

Emerson KJ, Merz CR, Catchen JM*, et al.* (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 16196-16200.

Glover KA, Skaala O, Limborg M, Kvamme C, Torstensen E (2011) Microsatellite DNA reveals population genetic differentiation among sprat (*Sprattus sprattus*) sampled throughout the Northeast Atlantic, including Norwegian fjords. *Ices Journal of Marine Science* **68**, 2145-2151.

Gnerre S, MacCallum I, Przybylski D*, et al.* (2011) High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 1513-1518.

Harnstrom K, Ellegaard M, Andersen TJ, Godhe A (2011) Hundred years of genetic structure in a sediment revived diatom population. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 4252-4257.

Herrera S, Reyes-Herrera PH, Shank TM (Chapter 2) Genome-wide predictability of restriction sites across the eukaryotic tree of life.

Herrera S, Shank TM (Chapter 5) RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa.

Hoegh-Guldberg O, Bruno JF (2010) The impact of climate change on the world's marine ecosystems. *Science* **328**, 1523-1528.

Hohenlohe PA, Bassham S, Etter PD*, et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* **6**, e1000862.

Johnstone IM (2001) On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics* **29**, 295-327.

Jones FC, Grabherr MG, Chan YF*, et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55-61.

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74**, 175-195.

Luikart G, Allendorf FW, Cornuet JM, Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity* **89**, 238-247.

Miller MR, Brunelli JP, Wheeler PA*, et al.* (2012) A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. *Molecular Ecology* **21**, 237-249.

Nedashkovskaya OI, Bum Kim S, Lysenko AM*, et al.* (2005) *Bizionia paragorgiae* gen. nov., sp. nov., a novel member of the family Flavobacteriaceae isolated from the soft coral *Paragorgia arborea*. *International Journal of Systematic and Evolutionary Microbiology* **55**, 375-378.

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genetics* **2**, e190.

Perrin C, Wing SR, Roy MS (2004) Effects of hydrographic barriers on population genetic structure of the sea star *Coscinasterias muricata* (Echinodermata, Asteroidea) in the New Zealand fjords. *Molecular Ecology* **13**, 2183-2195.

Prada C, Hellberg ME (2013) Long prereproductive selection and divergence by depth in a Caribbean candelabrum coral. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 3961-3966.

Price AL, Patterson NJ, Plenge RM*, et al.* (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**, 904-909.

Reitzel AM, Herrera S, Layden MJ, Martindale MQ, Shank TM (2013) Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular Ecology* **22**, 2953-2970.

Roberts JM, Wheeler A, Freiwald AR, Cairns SD (2009) *Cold-Water Corals : The Biology and Geology of Deep-Sea Coral Habitats* Cambridge University Press, Cambridge, UK ; New York.

Sánchez JA (2005) Systematics of the bubblegum corals (Paragorgiidae: Octocorallia: Cnidaria) with description of new species from New Zealand and the Eastern Pacific. *Zootaxa* **1014**, 1-72.

Suneetha KB, Salvanes AGV (2001) Population genetic structure of the glacier lanternfish, *Benthosema glaciale* (Myctophidae) in Norwegian waters. *Sarsia* **86**, 203-212.

Tracy CA, Widom H (1994) Level-spacing distributions and the airy kernel. *Communications in Mathematical Physics* **159**, 151-174.

Wating L, France SC, Pante E, Simpson A (2011) Biology of deep-water octocorals. *Advances in Marine Biology* **60**, 41-122.

Wildes SL, Vollenweider JJ, Nguyen HT, Guyon JR (2011) Genetic variation between outer-coastal and fjord populations of Pacific herring (*Clupea pallasii*) in the eastern Gulf of Alaska. *Fishery Bulletin* **109**, 382-393.

CHAPTER 7

Summary and conclusions

Deep-sea hydrothermal vents and deep-sea coral ecosystems are some of the most conspicuous biological hotspots in the deep-sea. These ecosystems face increasing threats caused by human activities, such as bottom trawling and deep-sea mining. Knowledge of conservation targets is fundamental for the implementation of efficient conservation strategies that help mitigate these threats. Such knowledge must include well-founded taxonomic inventories that allow us to identify species and ecosystems at risk, as well as an understanding of their relatedness, genetic variance, distribution, connectivity patterns, and adaptation potential. Nonetheless, gaining this knowledge in deep-sea ecosystems is difficult due to the extreme challenges of working in this environment, combined with the paucity of genetic resources for deep-sea organisms. Here I provide fundamental high-priority knowledge in taxonomic, evolutionary, and ecological aspects of deep-sea coral and vent species, by harnessing the power of novel genomic tools and overcoming long-standing methodological barriers.

In Chapter 2, I developed bioinformatic tools that help guide the design of studies aiming to characterize eukaryotic genome diversity using restriction-site associated DNA sequencing. With these tools I performed *in silico* genome-wide surveys thought the eukaryotic tree of life. I tested the hypothesis that genome composition, in terms of GC content, and mono-, di- and trinucleotide composition, can be used to predict the number of restriction sites for a given combination of restriction enzyme and genome across the eukaryotic tree of life. In most cases the trinucleotide genome composition model was the best predictor of the expected number of restriction sites in a eukaryotic genome, and the GC content and mononucleotide models the worst. I conclude that the predictability of restriction site frequencies in eukaryotic genomes needs to be treated on a case-specific basis, whereby the phylogenetic position of the taxon of interest and the specific recognition sequence of the selected restriction enzyme are the chief foci among the most determinant factors. The knowledge gained in this chapter, and the bioinformatic tools developed, was applied in all other subsequent chapters.

In Chapter 3, I tested global-scale historical biogeographic hypothesis of vent fauna using barnacles as model. I characterized the global genetic diversity of vent barnacles to infer their time and place of origin, mode of dispersal, and diversification throughout the world's vents. The approach was to target a suite of

multiple loci in samples representing seven out of the eight described genera. I also performed restriction-site associated DNA sequencing (RAD-seq) on individuals from each species. Phylogenetic inferences indicated that vent barnacles have colonized deep-sea hydrothermal vents at least twice in history. The late Mesozoic/Cenozoic was the time of colonization and radiation of barnacles in vent ecosystems. Further analyses suggested that the western Pacific was the place of origin of the major vent barnacle lineage, followed by circumglobal colonization eastward along the southern hemisphere during the Neogene. The inferred time of origin rejects previous hypotheses of antiquity of vent taxa. The timing and the mode of origin, radiation and dispersal are consistent with the inferences made for other deep-sea taxa, including non-vent species, and are correlated with the occurrence of major geological events and mass extinctions. Thus, I suggest that the geological processes and dispersal mechanisms discussed here can explain current distribution patterns of many other marine taxa and have played an important role shaping deep-sea faunal diversity.

In Chapter 4, I examined genetic diversity patterns in vent barnacles at a regional scale. To test the hypothesis that seamounts behave as isolated island-like systems, where population connectivity is limited and endemicity is promoted, I examined genome-wide RAD-seq data from three hydrothermal vent barnacle species. I compared the genetic diversity and population structuring patterns of barnacle populations from seamount and spreading ridges. Among the study populations I found patterns of population genetic structuring that do not conform to the predictions from the seamount endemicity hypothesis. The patterns of genetic variation among individuals collected from seamount and spreading ridges, separated horizontally by hundreds of kilometers and vertically by hundreds of meters, did not reject the null hypothesis of panmixia within each species. I found that these inferences are largely insensitive to the *de novo* assembly parameters used to identify loci from sequence reads. In conclusion, I suggest that the seamount endemicity hypothesis warrants further testing using high-resolution genetic markers in other vent organisms with differing life history strategies (e.g. brooders) that may limit their dispersal potential, as well as in non-vent organisms, which are not exposed to evolutionary pressures imposed by the dynamic nature of hydrothermal vent systems.

I then moved on to resolve long-standing questions regarding species definitions and relationships in deep-sea corals. In Chapter 5, I demonstrated the empirical utility of RAD-seq by unambiguously resolving phylogenetic relationships among recalcitrant octocoral taxa with divergences greater than 80 million years. I objectively inferred robust species boundaries in the genus *Paragorgia,* which contains some of the most important ecosystem engineers in the deep-sea, by testing alternative taxonomy-guided or unguided species delimitation hypotheses using the Bayes factors delimitation method (BFD*) with

genome-wide SNP data. I presented conclusive evidence rejecting the current morphological species delimitation model for the genus *Paragorgia* and indicating the presence of cryptic species boundaries associated with environmental variables. I argue that the suitability limits of RAD-seq for phylogenetic inferences in divergent taxa cannot be assessed in terms absolute time, but depend on taxon-specific factors such as mutation rate, generation time and effective population size. Classic morphological taxonomy can greatly benefit from integrative approaches that provide objective tests to species delimitation hypothesis.

Finally, in Chapter 6, I explored the adaptation potential of deep-sea coral species to environmental changes by examining a case of adaptation to shallow water from the deep-sea. Few populations of some deep-sea coral species can be found in shallow (< 45 m) high-latitude fjord environments where they experience significantly different environmental conditions than their deep relatives. Therefore, these shallow-water populations are believed to inhabit the extremes of the species' physiological tolerances and likely have developed adaptations that enable them to colonize these shallow-water environments. I aimed to identify potential genomic regions that have enabled the successful adaptation to shallow-water in the deep-sea octocoral species *Paragorgia stephencairnsi*. To characterize the genome-wide genetic diversity of populations of *P. stephencairnsi* found in shallow-water populations and compare it to the genetic diversity from deep-water populations, I performed high-resolution genome-wide scans of single nucleotide polymorphisms through RAD-seq. I found patterns of significant population genetic differentiation among the examined populations of *P. stephencairnsi*, which are consistent with the hypothesis that larvae from outer deep populations seeded shallow-water inner fjord populations. Furthermore, I find candidate positive-selection markers shared between parallel comparisons of shallow and deep populations, and thus identify them as likely candidate makers for genomic regions involved in adaptation to the shallow-water fjord environment.

Overall, the results from this thesis constitute critical baseline data with which to assess potential effects of anthropogenic disturbances on deep-sea ecosystems. The species delimitation frameworks here developed will enable rapid species assignments as deep-sea specimens from newly explored geographical regions become available. This thesis lays groundwork for describing the impacts of natural selection on deep-sea coral species in the face of environmental changes. The software here developed, and the resulting databases constitute a valuable reference resource that will help guide the choice of restriction enzyme for any study using RAD-seq or related methods.

I anticipate that the use of novel genomic tools to study deep-sea organisms will accelerate the pace of knowledge acquisition, and thus greatly enhance our understanding of deep-sea ecosystems, their evolution, and their role in the global ecosystems network. However, the speed of genomic data generation has now outpaced the development of analytical tools, and thus there is a great need of developing novel ways to make full use of the information contained in large genomic datasets. 'Omics' techniques promise a fast and direct route to move from descriptive studies in deep-sea organisms to process-oriented studies, which will allow use to understand the mechanisms that have allowed life to thrive in this extreme environment. Understanding these mechanisms can also lead to the development of a myriad of applications that can directly improve human's lives.